

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:34:37 ; Search time 12.23 Seconds
(Without alignments)
13,469 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	185	3	US-08-985-526-36
2	41	100.0	195	1	US-08-159-784-2
3	38	92.7	183	4	US-09-206-059-2
4	30	73.2	154	3	US-09-191-647-10
5	30	73.2	191	1	US-08-159-784-3
6	30	73.2	254	1	US-08-948-616-9
7	30	73.2	254	2	US-09-193-510-9
8	30	73.2	254	4	US-09-368-402-9
9	30	73.2	307	2	US-08-948-616-3
10	30	73.2	307	2	US-09-193-510-3
11	30	73.2	307	4	US-09-368-402-3
12	30	73.2	395	1	US-08-723-938-3
13	30	73.2	395	1	US-09-080-538-3
14	30	73.2	420	4	US-09-008-271A-4
15	30	73.2	420	4	US-08-974-691-8
16	30	73.2	470	2	US-08-724-394A-10
17	30	73.2	775	2	US-08-714-070A-1
18	29	70.7	113	2	US-08-466-860-8
19	29	70.7	113	3	US-08-472-040A-8
20	29	70.7	113	4	US-08-276-776-8
21	29	70.7	113	4	US-08-471-209-8
22	29	70.7	343	2	US-08-878-989-5
23	29	70.7	343	4	US-09-272-796-5
24	29	70.7	1503	4	US-08-976-255-14
25	28	68.3	318	4	US-09-387-574-2
26	28	68.3	353	3	US-08-984-288-2
27	28	68.3	353	4	US-09-224-426-4

28	28	68.3	353	4	US-09-478-601-4	Sequence 4, Appl
29	28	68.3	400	5	PCT-US95-16472-2	Sequence 2, Appl
30	28	68.3	402	3	US-08-602-809-2	Sequence 2, Appl
31	28	68.3	422	4	US-09-224-426-2	Sequence 2, Appl
32	28	68.3	422	4	US-09-478-601-2	Sequence 2, Appl
33	28	68.3	468	4	US-09-355-115-8	Sequence 8, Appl
34	28	68.3	516	2	US-09-019-201A-3	Sequence 3, Appl
35	28	68.3	751	3	US-08-946-026-6	Sequence 6, Appl
36	28	68.3	454	1	US-08-469-486-52	Sequence 52, Appl
37	28	68.3	454	2	US-08-469-658-52	Sequence 52, Appl
38	27	65.9	22	2	US-09-013-634-8	Sequence 8, Appl
39	27	65.9	337	2	US-09-013-634-2	Sequence 7, Appl
40	27	65.9	471	3	US-09-106-464-2	Sequence 2, Appl
41	27	65.9	956	2	US-08-897-443-3	Sequence 3, Appl
42	27	65.9	1090	3	US-08-307-896-3	Sequence 3, Appl
43	27	65.9	1090	3	US-08-726-214-4	Sequence 4, Appl
44	27	65.9	1090	5	PCT-US95-11808-3	Sequence 3, Appl
45	27	65.9	1235	1	US-08-118-101A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-985-526-36
Sequence 36, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
THERAPY
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
City: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
Zip: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorow Jr., Robert G
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 100.0%; Score 41; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 169 SYIVLCIE 176

RESULT 2
US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8996
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 100.0%; Score 41; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 179 SYVLICIE 186

RESULT 3
US-09-206-059-2
Sequence 2, Application US/09206059
Patent No. 6201104
GENERAL INFORMATION:
APPLICANT: Macdonald, Nicholas
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
FILE REFERENCE: 05213-0370
CURRENT APPLICATION NUMBER: US/09/206,059
CURRENT FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-09-206-059-2

Query Match 92.7%; Score 38; DB 4; Length 183;
Best Local Similarity 87.5%; Pred. No. 2.1;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 168 SYVLICIE 175

RESULT 4
US-09-191-647-10
Sequence 10, Application US/09191647
Patent No. 6046015
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/191,647
CURRENT FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/065,544
EARLIER FILING DATE: 1997-11-14
EARLIER APPLICATION NUMBER: 60/081,057
EARLIER FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 154
TYPE: PRT
ORGANISM: mouse
US-09-191-647-10

Query Match 73.2%; Score 30; DB 3; Length 154;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 44 SYVLICIE 51

RESULT 5
US-08-159-784-3
Sequence 3, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 191
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-3

Query Match 73.2%; Score 30; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IVLCIE 8
Db 177 IVLCIE 182

RESULT 6
US-08-948-616-9
Sequence 9, Application US/08948616
Patent No. 5840539
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1223894

US-08-948-616-9

Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
Db 69 SYVLCIE 76

RESULT 7
US-09-193-510-9
Sequence 9, Application US/09193510
Patent No. 5981226
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1223894
US-09-193-510-9

Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
Db 69 SYVLCIE 76

RESULT 8
US-09-368-402-9

Sequence 9, Application US/09368402
Patent No. 6200790
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,402
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/193,510
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1223894
US-09-368-402-9

Query Match 73.2% Score 30; DB 4; Length 254;
Best Local Similarity 62.5% Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYLVICIE 8
DB 69 SYMMLCTE 76

RESULT 9
US-08-948-616-3
Sequence 3, Application US/08948616
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-08-948-616-3

Query Match 73.2% Score 30; DB 2; Length 307;
Best Local Similarity 62.5% Pred. No. 1,2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYLVICIE 8
DB 69 SYMMLCTE 76

RESULT 10
US-09-193-510-3
Sequence 3, Application US/09193510
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-09-193-510-3

Query Match 73.2%; Score 30; DB 2; Length 307;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
11:1111
Db 69 SYMMLCTE 76

RESULT 11
US-09-368-402-3
Sequence 3, Application US/09368402
Patent No. 6200790
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,402
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/193,510
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-09-368-402-3

Query Match 73.2%; Score 30; DB 4; Length 307;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
11:1111
Db 69 SYMMLCTE 76

RESULT 12
US-08-723-938-3
Sequence 3, Application US/08723938
Patent No. 576759
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,938
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGTOT02
CLONE: 312099
US-08-723-938-3

Query Match 73.2%; Score 30; DB 1; Length 395;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YIVLCIE 8

Db 313 YIILCSE 319

RESULT 13

US-09-080-538-3
Sequence 3, Application US/09080538
Patent No. 5965129
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,538
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/723,938
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGNOT02
CLONE: 312099
US-09-080-538-3

Query Match 73.2% Score 30; DB 2; Length 395;
Best Local Similarity 71.4% Pred. NO. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YIILCIE 8
Db 313 YIILCSE 319

RESULT 14
US-09-008-271A-4
Sequence 4, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.

Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGAST01
CLONE: 877617
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-008-271A-4

Query Match 73.2% Score 30; DB 4; Length 420;
Best Local Similarity 71.4% Pred. NO. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YIILCIE 8
Db 313 YIILCSE 319

RESULT 15
US-08-974-691-8
Sequence 8, Application US/08974691
Patent No. 6225103
GENERAL INFORMATION:
APPLICANT: Keolsch, Gerald
APPLICANT: Lin, Xinli
APPLICANT: Tang, Jordan
TITLE OF INVENTION: Cloning and Characterization of Napsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 W. Peachtree
STREET: St.
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/974,691
: FILING DATE: 20-NOV-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/031,196
: FILING DATE: 20-NOV-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/046,126
: FILING DATE: 09-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Pabst, Patrea L.
: REGISTRATION NUMBER: 31,284
: REFERENCE/DOCKET NUMBER: OMR 166
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-873-8794
: TELEFAX: 404-873-8795
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 420 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: US-08-974-691-8

```

```

Query Match          73.2%; Score 30; DB 4; Length 420;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      2 YIVLCIE 8
       11:1111
Db      313 YIILCE 319

```

Search completed: August 16, 2001, 11:41:24
 Job time: 407 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:32:42 ; Search time 18.16 Seconds
(without alignments)
26.707 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: A_Geneseq_0601.*
 - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 21: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 23: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	8	AAV18407	C-terminus of anti
2	41	100.0	48	AA848608	Murine endostatin
3	41	100.0	50	AA835587	Antiangiogenic pen
4	41	100.0	184	AAV18409	Endostatin protein
5	41	100.0	184	AAV08689	Murine endostatin
6	41	100.0	184	AAV70258	Murine angiogenesis
7	41	100.0	184	AA849380	Murine endostatin
8	41	100.0	185	AAV06197	Anti-angiogenic en
9	41	100.0	191	AA828398	Murine endostatin
10	41	100.0	207	AA871930	Murine endostatin
11	41	100.0	218	AAV08691	Murine gene therap

12	41	100.0	580	20	AAV08692	Murine gene therap
13	41	100.0	684	20	AAV25114	Mouse alpha-1 (XVI)
14	41	100.0	1288	18	AAV26328	Mouse alpha-1 (XVI)
15	41	100.0	1288	18	AAV22297	Mouse alpha-1 (XVI)
16	38	92.7	27	22	AA849864	Human endostatin p
17	38	92.7	178	21	AAV94324	Human endostatin p
18	38	92.7	182	21	AA828399	Human endostatin p
19	38	92.7	182	21	AAV94323	Human endostatin p
20	38	92.7	182	21	AAV59622	Human endostatin p
21	38	92.7	183	20	AAV08693	Human endostatin p
22	38	92.7	183	20	AAV02113	Human endostatin p
23	38	92.7	183	21	AA830493	Human endostatin p
24	38	92.7	183	21	AA816451	Human endostatin p
25	38	92.7	183	21	AAV90771	Human endostatin p
26	38	92.7	183	21	AAV70252	Human endostatin p
27	38	92.7	183	22	AA849379	Human endostatin p
28	38	92.7	183	22	AA849810	Human endostatin p
29	38	92.7	193	21	AAW90877	Human endostatin p
30	38	92.7	195	21	AAW90874	Human endostatin p
31	38	92.7	216	21	AA830495	Human endostatin p
32	38	92.7	271	21	AA808407	Human endostatin p
33	38	92.7	684	18	AAW25113	Human endostatin p
34	38	92.7	684	18	AAW25113	Human endostatin p
35	38	92.7	1301	20	AAW92296	Human endostatin p
36	38	92.7	1336	20	AAW08694	Human endostatin p
37	36	87.8	356	22	AAW75233	Human endostatin p
38	36	87.8	184	21	AAV70265	Human endostatin p
39	33	80.5	184	22	AA849381	Human endostatin p
40	32	78.0	157	20	AAV37931	Human endostatin p
41	32	78.0	231	21	AA844187	Human endostatin p
42	32	78.0	260	21	AA844186	Human endostatin p
43	32	78.0	261	21	AA844185	Human endostatin p
44	31	75.6	112	20	AAV35789	Human endostatin p
45	31	75.6	113	21	AA856144	Human endostatin p

ALIGNMENTS

RESULT 1	
AAV18407	standard; peptide; 8 AA.
ID	AAV18407;
AC	AAV18407;
DT	24-AUG-1999 (first entry)
XX	
DE	C-terminus of anti-angiogenic peptide EMI.
KW	EMI: anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KW	benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KW	Ostler-Weber Syndrome; myocardial angiogenesis; angiodysplasia; cancer;
KW	plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KW	diagnosis graft vascular access stenosis; renal cancer; therapy.
XX	
OS	Mus sp.
PN	WO9929855-A1.
PD	17-JUN-1999.
XX	
PF	08-DEC-1998; 98WO-US26057.
XX	
PR	16-NOV-1998; 98US-0108536.
PR	08-DEC-1997; 97US-0067888.
PR	22-APR-1998; 98US-0082663.
XX	
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PI	Sukhatne VP;
XX	
DR	WPI, 1999-385604/32.
XX	

PT Mutant endostatin having anti-angiogenic activity

XX Claim 1: Page 71: 105pp: English.

CC This sequence represents the C-terminus from the mutant endostatin (EM)
 CC of the invention, which has anti-angiogenic activity, and is designated
 CC EMI. Compositions comprising EMI or fusion proteins comprising EMI, are
 CC useful for treating diseases characterised by angiogenic activity, such
 CC as angiogenesis-dependent cancers, benign tumours, rheumatoid arthritis,
 CC psoriasis, ocular angiogenesis, Osler-Webber Syndrome, myocardial
 CC angiogenesis, plaque neovascularisation, telangiectasia, haemophilic
 CC joints, angiodioma, wound granulation, intestinal adhesions,
 CC atherosclerosis, scleroderma, hypertrophic scars, cat scratch disease,
 CC helicobacter pylori ulcers, dialysis graft vascular access stenosis,
 CC contraception and obesity. In particular, the diseases treatable by EMI
 CC comprise cancer, especially renal cancer. The methods provide a means for
 CC introducing EMI into mammalian cells via gene therapy, for production of
 CC EMI via recombinant means, as well as recombinant production of the EMI
 CC protein. EMI performs as well or better than whole endostatin. Use
 CC of EMI is advantageous for treatment of angiogenic diseases in that
 CC increasingly smaller peptides are more potent on a weight basis, and may
 CC be able to better penetrate tissues.

CC Sequence 8 AA:

Query Match 100.0%; Score 41; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3 4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 DB 1 SYIVLCIE 8

RESULT 2

AAB49808
 ID AAB49808 standard; Protein: 48 AA.

AC AAB49808;

XX 02-MAR-2001 (first entry)

DE Murine endostatin peptide fragment SEQ ID NO: 21.

XX Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
 KW cancer; inflammation; angiogenesis-dependent disease.

XX Mus musculus.

OS WO200067771-A1.

XX 16-NOV-2000.

PD 02-MAY-2000; 2000WO-US12063.

XX 06-MAY-1999; 99US-0132907.

PR 14-JUL-1999; 99US-0353333.

XX (BURN-) BURNHAM INST.

XX VUori K;

XX WPI; 2001-040937/05.

XX Endostatin peptide comprising at least four endostatin amino acid
 PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy

XX Claim 10; Page 124-125; 146pp: English.

CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of

CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inappropriate invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.

CC Sequence 48 AA:

Query Match 100.0%; Score 41; DB 22; Length 48;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 DB 37 SYIVLCIE 44

RESULT 3

AAB35587
 ID AAB35587 standard; Peptide: 50 AA.

AC AAB35587;

XX 14-FEB-2001 (first entry)

DE Antiangiogenic pentacontapeptide IV.

XX Antiangiogenic; angiogenesis; cancer; endostatin.

XX Synthetic.

OS WO200063249-A1.

XX 26-OCT-2000.

PD 11-APR-2000; 2000WO-EP03236.

XX 15-APR-1999; 99IT-MI00777.

XX (UYMI-) UNIV MILANO.

PA (UYFI-) UNIV FIRENZE.

XX Chillemi F, Francescato P, Ziche M;

XX WPI; 2001-007005/01.

XX Polypeptides derived from endostatin exhibiting antiangiogenic activity
 PT useful for treatment of angiogenesis-dependent tumours

XX Claim 5; Page 18; 28pp: English.

CC The present invention describes a number of peptides derived from
 CC endostatin which exhibit antiangiogenic activity. These may be used in
 CC the treatment of cancer. The present sequence is one of the peptides of
 CC the invention.

XX Sequence 50 AA:

Query Match 100.0%; Score 41; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 0.26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 DB 34 SYIVLCIE 41

```

RESULT 4
AAV18409
ID AAV18409 standard; Protein: 184 AA.
XX
AC AAV18409;
XX
DT 24-AUG-1999 (first entry)
XX
DE Endostatin protein sequence.
XX
KM EMI; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KM benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KM Osler-Webber Syndrome; myocardial angiogenesis; angiodioma; cancer;
KM plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KM dialysis graft vascular access stenosis; renal cancer; therapy.
XX
XX Mus sp.
XX
XX WO929855-A1.
XX
XX 17-JUN-1999.
XX
XX 08-DEC-1998; 98WO-US26057.
XX
XX 16-NOV-1998; 98US-0108536.
XX
XX 08-DEC-1997; 97US-0067888.
XX
XX 22-APR-1998; 98US-0082663.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Sukhatme VP;
XX
XX WPI; 1999-385604/32.
XX
XX N-PSDB; AAX79949.
XX
XX Mutant endostatin having anti-angiogenic activity
XX
XX Claim 31; Fig 2; 105pp; English.
XX
XX This sequence is the mouse endostatin. The invention relates to a
XX the mutant endostatin (EMI), which has anti-angiogenic activity, and is
XX designated EMI. Compositions comprising EMI or fusion proteins comprising
XX EMI, are useful for treating diseases characterised by angiogenic
XX activity, such as angiogenesis-dependent cancers, benign tumours,
XX rheumatoid arthritis, psoriasis, ocular angiogenesis, Osler-Webber
XX Syndrome, myocardial angiogenesis, plaque neovascularisation,
XX telangiectasia, haemophilic joints, angiodioma, wound granulation,
XX intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,
XX cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular
XX access stenosis, contraception and obesity. In particular, the diseases
XX treatable by EMI comprise cancer, especially renal cancer. The methods
XX provide a means for introducing EMI into mammalian cells via gene
XX therapy, for production of EMI via recombinant means, as well as
XX recombinant production of the EMI protein. EMI performs as well or better
XX than whole endostatin. Use of EMI is advantageous for treatment of
XX angiogenic diseases in that increasingly smaller peptides are more potent
XX on a weight basis, and may be able to better penetrate tissues.
XX
XX Sequence 184 AA;

```

```

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYIVICIE 8
Db 168 SYIVICIE 175
RESULT 5
AAV08689

```

```

ID AAV08689 standard; Protein: 184 AA.
XX
AC AAV08689;
XX
XX 10-AUG-1999 (first entry)
XX
DE Murine endostatin protein fragment.
XX
XX Plasmidogen; murine; angiotensin; endostatin; gene therapy; vector;
XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
XX tumour growth; solid tumour; diabetic retinopathy; retina.
XX
XX Mus sp.
XX
XX WO926480-A1.
XX
XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-US24950.
XX
XX 20-NOV-1997; 97US-0975424.
XX
XX (GENE-) GENETIX PHARM INC.
XX (MAST) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Bachelot T, Leboulch P, Pawluc RJ;
XX
XX WPI; 1999-357696/30.
XX
XX N-PSDB; AAX77715.
XX
XX Anti-angiogenic gene therapy vectors
XX
XX Disclosure; Fig 6; 83pp; English.
XX
XX This invention describes a novel viral gene therapy vector comprising a
XX nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
XX from human or murine angiotensin, human or murine endostatin and
XX angiogenesis-inhibiting fusions and fragments, where the viral vector is
XX sufficiently attenuated for use in human gene therapy. The products of
XX the invention have anti-angiogenic, cytostatic, anti-diabetic and
XX ophthalmological activity. The vector is used in gene therapy for
XX inhibiting tumour growth in humans harbouring a solid tumour. The vector
XX expresses an anti-angiogenic polypeptide. An additional use comprises
XX treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
XX inhibits angiogenesis in the vicinity of the retina. The vector is
XX administered to cells ex vivo and then administered to the patient.
XX
XX Sequence 184 AA;

```

```

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYIVICIE 8
Db 168 SYIVICIE 175
RESULT 6
AAV70258
ID AAV70258 standard; Protein: 184 AA.
XX
AC AAV70258;
XX
XX 06-JUN-2000 (first entry)
XX
DE Murine angiogenesis inhibitor, endostatin.
XX
XX Murine; immunoglobulin Fc fragment; endostatin; immunofusion;
XX angiogenesis; inhibitor; cytostatic; antitumour; antiarthritic;
XX antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
XX vasotropic; vulnary; treatment; antiarteriosclerosis; tumour;

```

KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KM ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KM myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KM wound granulation; keloid scar; gene therapy.
 OS Mus musculus.
 XX WO200011033-A2.
 PN 02-MAR-2000.
 PD 02-MAR-2000.
 XX 25-AUG-1999; 99WO-US19329.
 PF 25-AUG-1999; 99WO-US19329.
 PR 25-AUG-1998; 98US-0097883.
 XX (LEXI-) LEXINGEN PHARM CORP.
 PA Lo K, Li Y, Gillies SD;
 PI WPI; 2000-237616/20.
 DR N-PSDB; AAZ51299.
 XX
 PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin
 PT FC region, useful for treating conditions mediated by angiogenesis,
 PT such as rheumatoid arthritis, tumors and macular degeneration -
 PS
 XX Example 5; Pages 48-49; 68pp; English.
 CC The patent discloses a DNA molecule encoding a fusion protein comprising
 CC a signal sequence, an immunoglobulin FC region, and an angiogenesis
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
 CC having angiostatin activity, a collagen XVIII fragment having endostatin
 CC activity, or combinations of them. The fusion protein (immunofusin) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumours,
 CC blood born tumours, tumour metastasis, benign tumours including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubeosis and Osler-Weber syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiodioma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
 CC in gene therapy. The present sequence is a murine
 CC endostatin used in the construction of immunofusin containing murine
 CC immunoglobulin FC fragment.
 CC
 SQ Sequence 184 AA:
 QY
 Db 168 sylvic16 175
 QY 1 SYIVICIE 8
 Db 168 sylvic16 175
 Query Match 100.0%; Score 41; DB 21; Length 184;
 Best Local Similarity 100.0%; Pred. No. 0.89; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 7
 AAB49380
 ID AAB49380 standard; Protein: 184 AA.
 XX
 AC AAB49380;
 XX
 DT 02-MAR-2001 (first entry)
 DE Murine endostatin SEQ ID NO: 4.
 XX
 DE Endostatin: antiangiogenic; angiogenesis; human; mouse; chicken;
 XX
 KW cancer; inflammation; angiogenesis-dependent disease.

XX
 OS Mus musculus.
 XX
 PN WO200067771-A1.
 XX
 PD 16-NOV-2000.
 XX
 PF 02-MAY-2000; 2000WO-US12063.
 PR 06-MAY-1999; 99US-0132907.
 PR 14-JUL-1999; 99US-0353333.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Wuori K;
 XX
 DR WPI; 2001-040937/05.
 DR N-PSDB; AAC88290.
 XX
 PT Endostatin peptide comprising at least four endostatin amino acid
 PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy -
 PS
 XX Disclosure; Fig 1; 146pp; English.
 CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of
 CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inappropriate invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.
 CC
 SQ Sequence 184 AA:
 QY
 Db 168 sylvic16 175
 QY 1 SYIVICIE 8
 Db 168 sylvic16 175
 Query Match 100.0%; Score 41; DB 22; Length 184;
 Best Local Similarity 100.0%; Pred. No. 0.89; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 8
 AAY06197
 ID AAY06197 standard; Protein: 185 AA.
 XX
 AC AAY06197;
 XX
 DT 16-AUG-1999 (first entry)
 DE Anti-angiogenic endostatin peptide.
 XX
 DE Anti-angiogenic endostatin peptide.
 XX
 KW Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;
 KW endostatin; melanoma; lung cancer; colon cancer; brain cancer;
 KW breast cancer.
 XX
 OS Homo sapiens.
 XX
 DE Key Location/Qualifiers
 FT Misc-difference 36 /note- "encoded by CAG"
 FT Misc-difference 37 /note- "encoded by CAA"
 FT Misc-difference 39 /note- "deduced sequence from nucleotide sequence

FT FT has an Ala residue between residues 39 and
 FT Misc-difference 76 /note- "encoded by AAC"
 FT Misc-difference 118 /note- "encoded by AAC"
 FT Misc-difference 162 /note- "encoded by AAA"
 FT Misc-difference 168 /note- "encoded by AAC"
 FT Misc-difference 185 /note- "encoded by AAA"
 FT Misc-difference 185 /note- "encoded by AAA"
 XX EP921193-A1.
 XX 09-JUN-1999.
 XX 07-JAN-1998; 98EP-0100135.
 XX 05-DEC-1997; 97US-0985526.
 XX (MIXS/) MIXSON A J.
 XX Mixson AJ;
 XX WPI; 1999-315406/27.
 XX N-PSDB; AAX58740.
 XX Inhibition of growth of solid tumors
 XX PS Disclosure; Page 38; 46pp; English.
 XX The present sequence represents an anti-angiogenic endostatin
 CC peptide. The invention provides a carrier:DNA complex that comprises
 CC DNA (see AAX58725-42) encoding an anti-angiogenic protein or peptide,
 CC such as the present sequence, the complex being deliverable to
 CC the site of a tumour in vivo, and additionally comprises regulatory
 CC elements for expressing the anti-angiogenic DNA in a tumour or
 CC tumour vasculature. The complex may also include DNA encoding a
 CC tumour suppressor protein, especially p53. The carrier is a
 CC liposome, cationic polymer, micelle, microsphere, virus, viral
 CC component, or a combination of these, and administration is by
 CC intravenous or intratumoral injection. The complexes are useful in
 CC gene therapy for inhibition of tumour growth. The types of tumors
 CC which may be treated include solid tumors such as melanomas and
 CC tumors in the lung, colon, brain and breast.
 XX SQ Sequence 185 AA;
 XX
 XX Query Match 100.0%; Score 41; DB 20; Length 185;
 XX Best Local Similarity 100.0%; Pred. No. 0.89;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1 SYIVLCIE 8
 XX |||||
 XX Db 169 syivlcie 176
 XX
 XX RESULT 9
 XX AAB28398
 XX ID AAB28398 standard; Protein: 191 AA.
 XX AAB28398;
 XX 19-FEB-2001 (first entry)
 XX Murine endostatin.
 XX Murine endostatin.
 XX Murine; endostatin; cytostatic; antiproliferative;
 XX vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
 XX cancer; vascularised solid tumour.

OS Mus sp.
 XX WO200064946-A2.
 XX 02-NOV-2000.
 XX 28-APR-2000; 2000WO-US11367.
 XX 28-APR-1999; 99US-0131432.
 XX (TEXA) UNITV TEXAS SYSTEM.
 XX Thorpe PE, Brecken RA;
 XX WPI; 2000-687317/67.
 XX N-PSDB; AAC67777.
 XX Immunogenic composition for the treatment and diagnosis of cancer
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
 XX Example 10; Page 290-291; 298pp; English.
 XX The present invention relates to anti-Vascular Endothelial Growth Factor
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
 CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
 CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.
 CC The anti-VEGF antibodies of the present invention are useful for the
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.
 XX SQ Sequence 191 AA;
 XX
 XX Query Match 100.0%; Score 41; DB 21; Length 191;
 XX Best Local Similarity 100.0%; Pred. No. 0.92;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1 SYIVLCIE 8
 XX |||||
 XX Db 175 syivlcie 182
 XX
 XX RESULT 10
 XX AAB71930
 XX ID AAB71930 standard; Protein: 207 AA.
 XX AAB71930;
 XX 10-MAY-2001 (first entry)
 XX Murine endostatin attached to Ig-kappa leader sequence.
 XX Mouse; endostatin; antitumour; cytostatic; angiogenic inhibitor;
 XX antidiabetic; ophthalmological; gene therapy; cardiovascular disease;
 XX adenoviral vector; diabetic retinopathy; cardiovascular disease;
 XX arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;
 XX lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.
 XX Mus sp.
 XX WO200112830-A1.
 XX 22-FEB-2001.
 XX 11-AUG-2000; 2000WO-EP07865.
 XX 13-AUG-1999; 99US-0373938.
 XX (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Hallenbeck PL, Chen CT;
XX
XX WPI: 2001-202871/20.
DR N-PSDB; AAF60336.
XX
PT Adenoviral vector for treating tumors and disorders associated with
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA
PT sequence encoding an angiogenic inhibitor, particularly endostatin
XX
PS Example 1; Fig 1B; 59pp; English.
XX
CC The nucleotide sequence encoding this protein was used in the
CC construction of an adenoviral vector which includes a DNA sequence
CC encoding endostatin. The adenoviral vector is useful for expressing
CC endostatin in a mammalian cell such as an A549 or Hep3b cell. It is
CC useful for treating other diseases and disorders associated with
CC angiogenesis, such as neovascular diseases and disorders associated with
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The
CC vector inhibits, prevents or destroys the growth of tumours by
CC preventing the formation of blood vessels in tumours, such as lymphoma
CC and leukaemia.
XX
SO Sequence 207 AA;

Query Match 100.0%; Score 41; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 191 sylvlicie 198
|||||||
RESULT 11
AA08691
ID AAY08691 standard; Protein; 218 AA.
XX
AC AAY08691;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-Flag-Endo.
XX
KW Plasmidogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO9926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawllok RJ;
XX
DR WPI: 1999-357696/30.
XX
DR N-PSDB; AAX77717.
XX
PT Anti-angiogenic gene therapy vectors
PS Example 1; Page 69; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a

CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SO Sequence 218 AA;

Query Match 100.0%; Score 41; DB 20; Length 218;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 202 sylvlicie 209
|||||||
RESULT 12
AA08692
ID AAY08692 standard; Protein; 580 AA.
XX
AC AAY08692;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.
XX
KW Plasmidogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO9926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawllok RJ;
XX
DR WPI: 1999-357696/30.
XX
DR N-PSDB; AAX77718.
XX
PT Anti-angiogenic gene therapy vectors
PS Example 1; Page 72-74; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.

XX
SQ Sequence 580 AA;

Query Match 100.0%; Score 41; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTVLCIE 8
Db 564 sytvlcie 571

RESULT 13

AAV25114
ID AAV25114 standard; Protein; 684 AA.

XX AAV25114;

XX 25-AUG-1999 (first entry)

DE Mouse alpha1 (XVIII) collagen protein.

XX Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KM anti-angiogenic; heparin binding domain; receptor binding domain; mimetic;
KM alpha-helix A domain; carbohydrate recognition domain; CRD domain;
KM treatment; angiogenesis; tumour; murine.

XX Mus sp.

XX WO9931616-A1.

XX 24-JUN-1999.

XX 16-DEC-1998; 98WO-0526783.

XX 16-DEC-1997; 97US-0069727.

XX (HARD) HARVARD COLLEGE.

XX Hohenester E, Olsen BR, Sasaki T, Timpl R;

XX WPI; 1999-395243/33.

XX Identifying mimetics of mammalian endostatin

XX Disclosure; Fig 5A-C; 75pp; English.

XX This invention describes a novel method for identifying mimetics of
XX mammalian endostatin. The method comprises identifying a compound
XX having atomic coordinates with non-trivial similarity to selected
XX coordinates of atoms of a mammalian endostatin involves (a) providing
XX a library of atomic coordinates of compounds in a library of candidate
XX compounds, (b) comparing the library of atomic coordinates to the
XX selected coordinates of a mammalian endostatin and (c) selecting from the
XX library at least one candidate compound on the basis of selection
XX criteria which include similarities between the atomic coordinates of the
XX selected candidate compound and the atomic coordinates of the mammalian
XX endostatin. The invention also describes the use of an anti-angiogenic
XX fragment of endostatin comprising a domain selected from a heparin
XX binding domain, a receptor binding domain, and exposed on alpha-helix A
XX domain, and a carbohydrate recognition domain (CRD) domain. The methods
XX can be used for designing and selecting endostatin mimics. The compounds
XX identified can be used for treating undesired angiogenesis, e.g. tumours.
XX This sequence represents mouse alpha1(XVIII) collagen which is used in
XX the description of the method.

XX Sequence 684 AA;

Query Match 100.0%; Score 41; DB 20; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTVLCIE 8
Db 669 sytvlcie 676

RESULT 14

AAW26328
ID AAW26328 standard; Protein; 1288 AA.

XX AAW26328;

XX 19-NOV-1997 (first entry)

DE Mouse alpha-1 collagen (XVIII).

XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.

XX Mus musculus.

XX Key

XX Location/Qualifiers

FT Peptide
FT /label= GXYGX'Y'_motif
FT 303..308
FT Peptide
FT /label= GXYGX'Y'_motif
FT 309..314
FT Peptide
FT /label= GXYGX'Y'_motif
FT 315..320
FT Peptide
FT /label= GXYGX'Y'_motif
FT 321..326
FT Peptide
FT /label= GXYGX'Y'_motif
FT 337..342
FT Peptide
FT /label= GXYGX'Y'_motif
FT 343..348
FT Peptide
FT /label= GXYGX'Y'_motif
FT 349..354
FT Peptide
FT /label= GXYGX'Y'_motif
FT 355..360
FT Peptide
FT /label= GXYGX'Y'_motif
FT 361..366
FT Peptide
FT /label= GXYGX'Y'_motif
FT 367..372
FT Peptide
FT /label= GXYGX'Y'_motif
FT 373..378
FT Peptide
FT /label= GXYGX'Y'_motif
FT 379..384
FT Peptide
FT /label= GXYGX'Y'_motif
FT 385..390
FT Peptide
FT /label= GXYGX'Y'_motif
FT 396..401
FT Peptide
FT /label= GXYGX'Y'_motif
FT 402..407
FT Peptide
FT /label= GXYGX'Y'_motif
FT 435..440
FT Peptide
FT /label= GXYGX'Y'_motif
FT 441..446
FT Peptide
FT /label= GXYGX'Y'_motif
FT 447..452
FT Peptide
FT /label= GXYGX'Y'_motif
FT 453..458
FT Peptide
FT /label= GXYGX'Y'_motif
FT 459..464
FT Peptide
FT /label= GXYGX'Y'_motif
FT 470..475
FT Peptide
FT /label= GXYGX'Y'_motif
FT 476..481
FT Peptide
FT /label= GXYGX'Y'_motif
FT 482..487
FT Peptide
FT /label= GXYGX'Y'_motif
FT 488..493
FT Peptide
FT /label= GXYGX'Y'_motif
FT 494..499
FT Peptide
FT /label= GXYGX'Y'_motif
FT 500..505
FT Peptide
FT /label= GXYGX'Y'_motif

```

FT Peptide 506..511 /label= GXYGX'Y'_motif
FT Peptide 512..517 /label= GXYGX'Y'_motif
FT Peptide 518..523 /label= GXYGX'Y'_motif
FT Peptide 524..529 /label= GXYGX'Y'_motif
FT Peptide 530..535 /label= GXYGX'Y'_motif
FT Peptide 536..541 /label= GXYGX'Y'_motif
FT Peptide 542..547 /label= GXYGX'Y'_motif
FT Peptide 548..553 /label= GXYGX'Y'_motif
FT Peptide 580..585 /label= GXYGX'Y'_motif
FT Peptide 586..591 /label= GXYGX'Y'_motif
FT Peptide 592..597 /label= GXYGX'Y'_motif
FT Peptide 598..603 /label= GXYGX'Y'_motif
FT Peptide 604..609 /label= GXYGX'Y'_motif
FT Peptide 610..615 /label= GXYGX'Y'_motif
FT Peptide 616..621 /label= GXYGX'Y'_motif
FT Peptide 622..627 /label= GXYGX'Y'_motif
FT Peptide 628..633 /label= GXYGX'Y'_motif
FT Peptide 634..639 /label= GXYGX'Y'_motif
FT Peptide 640..665 /label= GXYGX'Y'_motif
FT Peptide 657..662 /label= GXYGX'Y'_motif
FT Peptide 677..682 /label= GXYGX'Y'_motif
FT Peptide 683..688 /label= GXYGX'Y'_motif
FT Peptide 689..694 /label= GXYGX'Y'_motif
FT Peptide 695..700 /label= GXYGX'Y'_motif
FT Peptide 707..712 /label= GXYGX'Y'_motif
FT Peptide 713..718 /label= GXYGX'Y'_motif
FT Peptide 735..740 /label= GXYGX'Y'_motif
FT Peptide 741..746 /label= GXYGX'Y'_motif
FT Peptide 747..752 /label= GXYGX'Y'_motif
FT Peptide 759..764 /label= GXYGX'Y'_motif
FT Peptide 765..770 /label= GXYGX'Y'_motif
FT Peptide 771..776 /label= GXYGX'Y'_motif
FT Peptide 787..792 /label= GXYGX'Y'_motif
FT Peptide 793..798 /label= GXYGX'Y'_motif
FT Peptide 799..804 /label= GXYGX'Y'_motif
FT Peptide 815..820 /label= GXYGX'Y'_motif
FT Peptide 821..826 /label= GXYGX'Y'_motif

```

```

FT Peptide /label= GXYGX'Y'_motif
FT Peptide 827..832 /label= GXYGX'Y'_motif
FT Peptide 833..838 /label= GXYGX'Y'_motif
FT Peptide 839..844 /label= GXYGX'Y'_motif
FT Peptide 845..850 /label= GXYGX'Y'_motif
FT Peptide 863..868 /label= GXYGX'Y'_motif
FT Peptide 869..874 /label= GXYGX'Y'_motif
FT Peptide 875..880 /label= GXYGX'Y'_motif
FT Peptide 891..896 /label= GXYGX'Y'_motif
FT Peptide 897..902 /label= GXYGX'Y'_motif
FT Peptide 903..908 /label= GXYGX'Y'_motif
FT Peptide 911..916 /label= GXYGX'Y'_motif
FT Peptide 917..922 /label= GXYGX'Y'_motif
FT Peptide 928..933 /label= GXYGX'Y'_motif
FT Peptide 934..939 /label= GXYGX'Y'_motif
FT Peptide 956..961 /label= GXYGX'Y'_motif
FT Peptide 962..967 /label= GXYGX'Y'_motif
FT Peptide 968..973 /label= GXYGX'Y'_motif
FT Peptide 1126..1131 /label= GXYGX'Y'_motif
FT Peptide 1145..1150 /label= GXYGX'Y'_motif
FT Peptide 1193..1198 /label= GXYGX'Y'_motif
FT Peptide /label= GXYGX'Y'_motif

```

```

XX US5643783-A.
XX PN
XX PD 01-JUL-1997.
XX PF 01-DEC-1993; 93US-0159784.
XX PR 01-DEC-1993; 93US-0159784.
XX PA (HARD ) HARVARD COLLEGE.
XX PI Oh SP, Olsen BR:
XX DR WPI: 1997-350247/32.
XX DR N-PSDB: AAT84485.
XX Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX PS Disclosure; Fig 2; 35pp; English.

```

Query Match 100.0%; Score 41; DB 18; Length 1288;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYIVLCIE 8
 |||||
 Db 1272 syivlcie 1279

RESULT 15


```

AAW92297
ID AAW92297 standard; peptide; 1288 AA.
XX
AC AAW92297;
XX
DT 28-APR-1999 (first entry)
XX
DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
XX
KW Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW hepatocellular carcinoma; diagnosis.
XX
OS Mus sp.
XX
PN MO9856399-A1.
XX
PD 17-DEC-1998.
XX
PF 12-JUN-1998; 98MO-US12327.
XX
PR 12-JUN-1997; 97US-0049369.
XX
PA (FIFT-) ACAD FINLAND.
PA (FIBR-) FIBROGEN INC.
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI Clement B, Pihlajaniemi T, Rehn M;
XX
DR WPI; 1999-070292/06.
XX
PT Diagnosis and monitoring of liver disease by measuring collagen type
PT XVIII levels - with elevated levels indicative of disease,
PT especially cirrhosis or hepatocellular carcinoma
XX
PS Example 6; Fig 8; 56pp; English.
XX
CC A method has been developed for the detecting liver disease. The method
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
CC complex (C) formed as indicator of the amount of Coll18 present; (c)
CC similar analysis of a non-diseased control; and (d) comparing the
CC amounts of Coll18 in the two samples to detect presence or progression of
CC disease. Elevated levels of Coll18 are: (i) indicative of disease,
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,
CC specifically hepatocellular carcinoma (there is a relationship between
CC Coll18 mRNA levels and tumour size and necrosis, and survival times are
CC significantly higher in patients with higher Coll18 levels). The method
CC provides non-invasive, early and accurate diagnosis of liver disease.
CC The present sequence represents the sequence common to mouse alpha-1
CC (XVIII) collagen chain from the present invention.
XX
SQ Sequence 1288 AA;

```

```

Query Match 100.0%; Score 41; DB 20; Length 1288;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYIVICIE 8
Db 1273 syivicle 1280

```

Search completed: August 16, 2001, 11:41:06
Job time: 504 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:41:07 ; Search time 21.98 Seconds
(without alignments)
48.155 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175
Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP_mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	171	11	09WUW5
2	41	100.0	184	11	09UK63
3	41	100.0	226	11	09QK63
4	41	100.0	1140	11	09QK63
5	41	100.0	1774	11	061434
6	33	80.5	237	2	087543
7	33	80.5	386	13	093419
8	33	80.5	646	5	015900
9	33	80.5	706	5	015901
10	33	80.5	1905	5	09XTP6
11	32	78.0	147	2	084569
12	32	78.0	147	2	09PUS5
13	32	78.0	347	1	09P9G6
14	32	78.0	507	3	09P9G5
15	32	78.0	613	4	095942
16	32	78.0	1077	13	091019
17	32	78.0	1142	13	002391
18	32	78.0	1160	11	0921E9
19	32	78.0	1171	11	062638

20	32	78.0	1175	11	061543	061543 mus musculu
21	32	78.0	1177	4	013221	013221 homo sapien
22	32	78.0	1179	4	092896	092896 homo sapien
23	31	75.6	70	4	09PIN2	09PIN2 homo sapien
24	31	75.6	158	2	092784	092784 chlamydia p
25	31	75.6	163	10	09SLC0	09SLC0 arabidopsis
26	31	75.6	213	14	09J5B0	09J5B0 fowlpox vir
27	31	75.6	235	5	093865	093865 caenorhabd1
28	31	75.6	272	14	09BN36	09BN36 anisaceta moo
29	31	75.6	316	2	09RST2	09RST2 dehnococcus
30	31	75.6	466	2	09KMS0	09KMS0 vibrio chol
31	31	75.6	472	5	09VON6	09VON6 drosophila
32	31	75.6	479	10	09XIS2	09XIS2 arabidopsis
33	31	75.6	566	2	09PHU2	09PHU2 campylobact
34	31	75.6	991	5	095024	095024 dictyostell
35	31	75.6	2158	10	09LUT5	09LUT5 arabidopsis
36	31	75.6	2342	5	001677	001677 bombay mori
37	30	73.2	96	9	048440	048440 bacterioph
38	30	73.2	129	5	062211	062211 caenorhabd1
39	30	73.2	141	5	096254	096254 plasmodium
40	30	73.2	153	14	093567	093567 fowlpox vir
41	30	73.2	163	5	061832	061832 caenorhabd1
42	30	73.2	168	2	031938	031938 bacillus su
43	30	73.2	168	9	064083	064083 bacterioph
44	30	73.2	254	11	062891	062891 rattus norv
45	30	73.2	267	2	091060	091060 pseudomonas

ALIGNMENTS

RESULT 1
ID 09WUW5 PRELIMINARY; PRT: 171 AA.
AC 09WUW5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE COLLAGEN TYPE XVIII, ALPHA (1) CHAIN (FRAGMENT).
GN COL18A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Eberspacher U., Donner P., Schuppan D.;
RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;
RT "Temporopatial expression of collagen XVIII/endostatin in acute and
chronic liver injuries.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ236873; CAB44263.1; -;
DR HSSP: P39061; IMOE.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 171 171
SO SEQUENCE 171 AA; 18933 MW; 81BF2EB3FC2C8E72 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
DB 160 SYIVLCIE 167

```

RESULT 2
Q9JK63 PRELIMINARY: PRT: 184 AA.
AC Q9JK63;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ENDOSTATIN (FRAGMENT).
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE KUNMING;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor
neovasculature."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257775; AAF69009.1; -.
FT NON_TER 1
SQ SEQUENCE 184 AA: 20376 MW: AC06F9D8D103412A CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SYVLCE 8
Db 168 SYVLCE 175

```

```

RESULT 3
Q9QZD2 PRELIMINARY: PRT: 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Chen L., Perletti G., Folkman J.;
RT "Antitumor activity of rat endostatin."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF189709; AAF00975.1; -.
FT HSSP; P39061; IKOE.
KM collagen.
RT chain.
RN NON_TER 1
SQ SEQUENCE 226 AA: 25350 MW: 38B83C0486C0E949 CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SYVLCE 8
Db 210 SYVLCE 217

```

```

RESULT 4
Q61434 PRELIMINARY: PRT: 1140 AA.
ID Q61434;
AC Q61434;

```

```

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update).
DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94059075; PubMed=8240330;
RA Abe N., Muragaki Y., Yoshida H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196; 196:576-582(1993).
DR EMBL; D17546; BAA04483.1; -.
DR HSSP; P39061; IKOE.
DR MGD; MGI:88449; COL15A1.
DR InterPro; IPR000087; -.
DR Pfam; PF01391; Collagen; 6.
FT NON_TER 1
SQ SEQUENCE 1140 AA: 115156 MW: 8B0C7E6862B3BDFE CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SYVLCE 8
Db 1124 SYVLCE 1131

```

```

RESULT 5
Q62001 PRELIMINARY: PRT: 1774 AA.
ID Q62001;
AC Q62001; Q60672;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN
DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).
GN COL18A1.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hinkikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
partial structure of the corresponding gene, and comparison of the
alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
chain."
RL J. Biol. Chem. 269; 13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Pihlajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
collagenous sequence, a distinct tissue distribution, and homology
with type XV collagen."
RL Proc. Natl. Acad. Sci. U.S.A. 91; 4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Pihlajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
and tissue-specific differences in the expression of the corresponding
transcripts. The longest form contains a novel motif homologous to rat

```

RT and Drosophila frizzled proteins.";
 RL J. Biol. Chem. 270:4705-4711(1995).
 DR EMBL: U03715; AAC52903.1; -.
 DR EMBL: U03716; AAC52903.1; JOINED.
 DR EMBL: U03718; AAC52903.1; JOINED.
 DR EMBL: U34607; AAC52903.1; JOINED.
 DR EMBL: U34608; AAC52903.1; JOINED.
 DR EMBL: U34609; AAC52903.1; JOINED.
 DR EMBL: U34610; AAC52903.1; JOINED.
 DR EMBL: U34611; AAC52903.1; JOINED.
 DR EMBL: U34612; AAC52903.1; JOINED.
 DR EMBL: U34613; AAC52903.1; JOINED.
 DR EMBL: U11637; AAC52179.1; -.
 DR HSSP: P39061; IKOE.
 DR MGD: MGI:88451; Coll18a1.
 DR InterPro: IPR000024; -.
 DR InterPro: IPR001087; -.
 DR InterPro: IPR001791; -.
 DR InterPro: IPR003129; -.
 DR Pfam: PF01391; Collagen; 6.
 DR Pfam: PF01392; Fz; 1.
 DR Pfam: PF02210; TSPN; 1.
 DR SMART: SM00282; LAMG; 1.
 KW Signal.
 SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 1774;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYIVLCIE 8
 |||||
 Db 1758 SYIVLCIE 1765

RESULT 6
 ID 087543 PRELIMINARY; PRT: 237 AA.
 AC 087543;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE OUTER MEMBRANE PROTEIN ADHESIN (FRAGMENT).
 OS Capnocytophaga gingivalis.
 OC Bacteria; CPB group; Flavobacteriia; Flavobacteriaceae; Capnocytophaga.
 OX NCBI_TaxID=1017;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DR2001;
 RA Tempio P.J.;
 RT "Capnocytophaga outer-membrane adhesin DNA sequence, 3'end.";
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083393; AAC33553.1; -.
 DR MEROPS; 509.013; -.
 DR InterPro: IPR000379; -.
 DR InterPro: IPR001375; -.
 DR InterPro: IPR002471; -.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 FT NON_TER
 SQ SEQUENCE 237 AA; 27482 MW; 4D630AC81D64EDAC CRC64;

Query Match 80.5%; Score 33; DB 2; Length 237;
 Best Local Similarity 71.4%; Pred. No. 23;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 YIVLCIE 8
 |||||
 Db 62 YIVLCVD 68

RESULT 7
 ID 093419 PRELIMINARY; PRT: 386 AA.
 AC 093419;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
 DE COLLAGEN XVIII (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Halfter W., Dong S., Schurer B., Cole G.;
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
 RL J. Biol. Chem. 0:0-0(1998).
 DR EMBL: AF083440; AAC33294.1; -.
 DR HSSP: P39061; IKOE.
 DR InterPro: IPR000087; -.
 FT NON_TER
 SQ SEQUENCE 386 AA; 41775 MW; 34D40FA09EBA3B0E CRC64;

Query Match 80.5%; Score 33; DB 13; Length 386;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYIVLCIE 8
 ::|||
 Db 370 AFVLCIE 377

RESULT 8
 ID 015900 PRELIMINARY; PRT: 646 AA.
 AC 015900;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE PUTATIVE TRANSCRIPTION INHIBITOR.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.
 OX NCBI_TaxID=44889;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-TD-4;
 RA Wells D.J., Welker D.L.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U57081; AAB95435.1; -.
 DR InterPro: IPR001584; -.
 DR Pfam: PF00665; IVE; 1.
 SQ SEQUENCE 646 AA; 74364 MW; 89EAF091505B2D4E CRC64;

Query Match 80.5%; Score 33; DB 5; Length 646;
 Best Local Similarity 71.4%; Pred. No. 55;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 YIVLCIE 8
 |||||
 Db 185 YIICID 191

RESULT 9
 ID 015901 PRELIMINARY; PRT: 706 AA.
 AC 015901;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE PUTATIVE TRANSPORASE.

OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSDON-TD-4:
 RA Wells D.J., Welker D.L.
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 057081; AAB95436.1; -
 DR InterPro: IPR001584; -
 DR Pfam: PF00665; rve; 1.
 SQ SEQUENCE 706 AA; 80670 MW; F5718916484A3E5 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 706;
 Best Local Similarity 71.4%; Pred. No. 59;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 YVLCIE 8
 DB 185 YVLCID 191

RESULT 10
 ID Q9XTP6 PRELIMINARY; PRT; 1905 AA.
 AC Q9XTP6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CTRP PROTEIN PRECURSOR (OOKINERE PROTEIN).
 GN CTRP.
 OS Plasmodium berghei.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M., Sawal T., Chinzel Y.;
 RL "Structure and expression of an adhesive protein-like molecule of mosquito invasive-stage malaria parasite."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M., Sawal T., Chinzel Y.;
 RL "Structure and expression of an adhesive protein-like molecule of mosquito invasive-stage malaria parasite."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M.;
 RL "Structure and Expression of an Adhesive Protein-like Molecule of Mosquito Invasive-stage Malarial Parasite.";

RT J. Exp. Med. 0:0-0(1999).
 RL EMBL: A238798; CAB45562.1; -
 DR EMBL: AB027129; BA82322.1; -
 DR EMBL: AF149771; AAF73158.1; -
 DR InterPro: IPR000884; -
 DR InterPro: IPR002035; -
 DR Pfam: PF00090; tgp_1; 4.
 DR Pfam: PF00092; vva; 5.
 DR PRINTS: PRO0453; VWFADOMAIN.
 KW SIGNAL.
 FT CHAIN 22 1905 POTENTIAL.
 SQ SEQUENCE 1905 AA; 215138 MW; 27A94B778CABDC36 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 1905;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYVLCI 7
 DB 10 SYVLCV 16

RESULT 11
 ID 084569 PRELIMINARY; PRT; 147 AA.
 AC 084569;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 16.0 KDA PROTEIN.
 GN CT565.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/OW-3/CX;
 RX MEDLINE-99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 DR EMBL: AE001327; AAC68167.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 147 AA; 15980 MW; 19D0E2CE2F9DE4E CRC64;

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYVLCI 7
 DB 123 SYVLCI 129

RESULT 12
 ID Q9PJH5 PRELIMINARY; PRT; 147 AA.
 AC Q9PJH5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL PROTEIN TC0854.
 GN TC0854.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGS;
 RX MEDLINE-20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Ulteback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Koloney J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002352; AAF39650.1; -
 DR TIGR: TC0854; -
 KW Hypothetical protein.
 SQ SEQUENCE 147 AA; 15871 MW; 6B33E9C475F8FD66 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 147;

Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
DB 123 SYIVLCI 129

RESULT 13
Q9P9G6 PRELIMINARY; PRT; 347 AA.
AC 09P9G6;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 39.0 KDA PROTEIN.
OS Methanococcoides burtonii.
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanococcoides.
OX NCBI_TaxID=29291;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 6242;
RX MEDLINE=20198308; PubMed=10731411;
RA Lim J., Thomas T., Cavichioni R.;
RT "Low temperature regulated DEAD-box RNA helicase from the Antarctic
archaeon, Methanococcoides burtonii.";
RL J. Mol. Biol. 297:553-567(2000).
DR EMBL: AF199442; AAF89100.1; -;
KW Hypothetical protein.
SQ SEQUENCE 347 AA; 39010 MW; EEF82AC2CD1F6DC8 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 347;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
DB 196 SYIVLCI 202

RESULT 14
Q9P905 PRELIMINARY; PRT; 507 AA.
AC 09P905;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE PALC.
GN PALC.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Negrete-Urtasun S., Reiter W., Diez E., Denison S.H., Tiltburn J.,
RA Espeso E.A., Penabaz M.A., Arst H.N., Jr.;
RT "Ambient pH signal transduction in Aspergillus: completion of gene
characterization.";
RL Mol. Microbiol. 0:0-0(1999).
DR EMBL: AF152925; AAF70857.1; -;
SQ SEQUENCE 507 AA; 54580 MW; 3AA315413977F929 CRC64;

Query Match 78.0%; Score 32; DB 3; Length 507;
Best Local Similarity 62.5%; Pred. No. 72;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCI 8
DB 236 SYIVLCI 243

RESULT 15
O95942 PRELIMINARY; PRT; 613 AA.
AC O95942;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE ALPHA INTEGRIN INTERACTING PROTEIN 63 (FRAGMENT).
GN AIBP63.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Wixler V.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-235 FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99192311; PubMed=10094488;
RA Wixler V., Lapanline E., Geerts D., Sonnenberg A., Petersohn D.,
RA Eekes B., Paulsson M., Ammalle M.;
RT "Identification of novel interaction partners for the conserved
membrane proximal region of alpha-integrin cytoplasmic domains.";
RL PNAS Lett. 445:351-355(1999).
DR EMBL: AJ131721; CAB38232.1; -;
KW Integrin.
FT NON_TER
SQ SEQUENCE 613 AA; 69523 MW; F742491675A63F55 CRC64;

Query Match 78.0%; Score 32; DB 4; Length 613;
Best Local Similarity 57.1%; Pred. No. 85;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
DB 473 YIVLCIE 479

Search completed: August 16, 2001, 11:44:01
Job time: 174 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:41:28 ; Search time 10.05 Seconds
(without alignments)
27.268 Million cell updates/sec

Title: US-09-589-777a-2_COPY_168_175

Perfect score: 41
Sequence: 1 SYVLICIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwisProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1527	1 CAIH_MOUSE	P39061 mus musculus
2	38	92.7	1516	1 CAIH_HUMAN	P39060 homo sapien
3	32	78.0	1015	1 YA17_HUMAN	O94923 homo sapien
4	30	73.2	233	1 YAS6_MEIJA	O58442 methanococ
5	30	73.2	346	1 RDS_FEICA	P35906 felis silve
6	30	73.2	407	1 REC_SALTY	P26479 salmoneila
7	30	73.2	420	1 NAPI_HUMAN	O96009 homo sapien
8	30	73.2	436	1 NPT3_HUMAN	O00624 homo sapien
9	30	73.2	550	1 YW40_YEAST	O03212 saccharomyc
10	30	73.2	769	1 COMP_BACSU	O99027 bacillus su
11	30	73.2	775	1 TH11_SCHPO	P36598 schizosacch
12	30	73.2	856	1 VPH1_NEUCR	O01290 neurospora
13	30	73.2	880	1 CAIE_HUMAN	P39059 homo sapien
14	29	70.7	1388	1 YOT0_CAEEL	P34655 caenorhabdi
15	29	70.7	325	1 MCSR_HUMAN	P33032 homo sapien
16	29	70.7	487	1 MOR4_HUMAN	O15374 homo sapien
17	29	70.7	602	1 CRK_DAUCA	P53681 daucus caro
18	29	70.7	891	1 YB33_SCHPO	O14338 schizosacch
19	29	70.7	923	1 YAVA_SCHPO	O10165 schizosacch
20	29	70.7	4543	1 LRPI_CHICK	P08157 gallus gall
21	28	68.3	79	1 PSPB_PIG	P15782 sus scrofa
22	28	68.3	125	1 YN25_CAEEL	P31584 caenorhabdi
23	28	68.3	146	1 HBB_MEGLY	P11752 megaderma 1
24	28	68.3	157	1 Y173_METVA	O57637 methanococ
25	28	68.3	171	1 IRI0_HCMVA	P16808 human cytom
26	28	68.3	178	1 LACB_PIG	P04119 sus scrofa
27	28	68.3	213	1 AT12_VACCV	P24758 vaccinia vi
28	28	68.3	236	1 VHEL_MCMVM	P08499 white clove
29	28	68.3	253	1 TPIS_BORBU	P14403 white clove
30	28	68.3	255	1 CTRC_NEIMB	O59182 borrelia bu
31	28	68.3	294	1 YFEC_YEYPE	P32015 neisseria m
32	28	68.3	305	1 CAG7_RAT	O56954 yersinia pe
33	28	68.3	305	1 CAG7_RAT	O64686 rattus norv

34	28	68.3	322	1 AT12_VACCC	P21114 vaccinia v1
35	28	68.3	325	1 MCSR_BOVIN	P56451 bos taurus
36	28	68.3	340	1 YDDR_ECOLI	P77308 escherichia
37	28	68.3	353	1 GPRO_RAT	P97639 rattus norv
38	28	68.3	381	1 YCU4_SCHPO	O74405 schizosacch
39	28	68.3	402	1 GPRO_HUMAN	O99705 homo sapien
40	28	68.3	435	1 U1A9_HSVSA	O01046 herpesvirus
41	28	68.3	496	1 YKAB_CAEEL	P34262 caenorhabdi
42	28	68.3	525	1 SYH_CAEEL	P34183 caenorhabdi
43	28	68.3	540	1 YFEO_YEAST	P43562 saccharomyc
44	28	68.3	560	1 FOBI_YEAST	O13329 saccharomyc
45	28	68.3	596	1 AGP2_YEAST	P38090 saccharomyc

ALIGNMENTS

RESULT 1
CAIH_MOUSE STANDARD: PRT: 1527 AA.
ID CAIH_MOUSE Q62002; Q61437;
AC P39061; Q62002; Q61437;
DT 01-FEB-1995 (Rel. 31, Created)
DR 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR [CONSTAINS: ENDOSTATIN].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN-BALB/C; TISSUE-Liver;
RX MEDLINE-94245707; PubMed-8188673;
RA Reim M.V., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.";
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RA Reim M., Hintikka E., Pihlajaniemi T.;
RT "Characterization of the mouse gene for the alpha-1 chain of type
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RX MEDLINE-94240112; PubMed-8183894;
RA Reim M.V., Pihlajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [4]
RP SEQUENCE OF 240-1527 FROM N.A.
RX TISSUE-Liver;
RC MEDLINE-94240111; PubMed-8183893;
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT "Isolation and sequencing of cDNAs for proteins with multiple domains
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN [5]
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE-94240118; PubMed-9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth.";
RL Cell 88:277-285(1997).

[illegible]

FT	DOMAIN	1087	1098	NONHELICAL REGION 8 (NC8).
FT	DOMAIN	1099	1122	TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	1123	1129	NONHELICAL REGION 9 (NC9).
FT	DOMAIN	1130	1161	TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	1162	1194	NONHELICAL REGION 10 (NC10).
FT	DOMAIN	1195	1212	TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN	1213	1527	NONHELICAL REGION 11 (NC11).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	700	700	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	1376	1516	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	1478	1508	
FT	SITE	1104	1106	CELL ATTACHMENT SITE (POTENTIAL).
FT	VARSPLIC	1	212	MISSING (IN SHORT ISOMER).
FT	VARSPLIC	213	238	AVTQTPPPROSNQIAVLPSPAPDF -> MAPRMHLDVLT
FT	CONFLICT	900	900	ISVLLAVRVSAWAE (IN SHORT ISOMER).
FT	CONFLICT	947	947	P -> L (IN REF. 4).
FT	CONFLICT	964	964	A -> F (IN REF. 4).
FT	CONFLICT	1157	1157	R -> P (IN REF. 4).
FT	CONFLICT	1266	1266	P -> L (IN REF. 4).
FT	CONFLICT	1276	1276	L -> F (IN REF. 4).
FT	CONFLICT	1437	1437	L -> V (IN REF. 4).
SQ	SEQUENCE	1527	AA: 156008	MM: 9645045AF140B513 CRC64:

```

Query March      100.0%; Score 41; DB 1; Length 1527;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
Oy      1 SYVICIE 8
          |||||
Db      1511 SYVICIE 1518

```

RESULT	2
CAIH_HUMAN	ID
CAIH_HUMAN	STANDARD;
P39060; Q9Y608; Q9Y607; Q9UC38;	PRT; 1516 AA.
01-FEB-1995 (Rel. 31, Created)	
01-OCT-2000 (Rel. 40, Last sequence update)	
01-OCT-2000 (Rel. 40, Last annotation update)	
COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].	
COL18A1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
NCBI_TaxID=9606;	
[1]	
SEQUENCE FROM N.A.	
RX MEDLINE-98164096; PubMed=9503365;	
RX Saetela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;	
RX "Complete primary structure of two variant forms of human type XVIII	
RX collagen and tissue-specific differences in the expression of the	
RX corresponding transcripts [1998].	
RL Matrix Biol. 16:319-328(1998).	
RN [2]	
RN SEQUENCE FROM N.A.	
RX MEDLINE-20289799; PubMed=10830953;	
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,	
RX Park H.-S., Toyoda A., Ishii K., Tokioki Y., Choi D.-K., Soeda F.,	
RX Ohki M., Takagi T., Sakaki K., Lehmann R., Patterson D.,	
RX Menzel U., Delabard T., Kumpf K., Lehmann R., Patterson D.,	
RX Reichwald A., Rump A., Schillhabel M., Schudy A., Zimmermann W.,	
RX Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Aakawa S.,	
RX Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,	
RX Minoshima S., Shimizu N., Nordstok G., Hornischer K., Brandt P.,	
RX Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,	
RX Ransier J., Beck A., Klages S., Hennig S., Resselmann L., Dagand E.,	
RX Mehrezyer S., Borzjz M., Gaudier K., Nizetic D., Francis F.,	
RX Lehrsich R., Reinhardt R., Yaspo M.-L.;	
RT "The DNA sequence of human chromosome 21.;"	
RL Mature 405:311-319(2000).	
RN [3]	

RP SEQUENCE OF 834-1516 FROM N.A.
 RX MEDLINE-94245237; PubMed-8186291;
 RA Oh S.P., Wairman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
 RA Olsen B.R.;
 RT *Cloning of cDNA and genomic DNA encoding human type XVIII collagen
 RT and localization of the alpha 1(XVIII) collagen gene to mouse
 RT chromosome 10 and human chromosome 21.*;
 RT Genomics 19:494-499(1994).
 RL [4]
 RN SEQUENCE OF 1334-1516 FROM N.A.
 RP TISSUE-Placenta;
 RC Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
 RA *Cloning and expression of human endostatin gene in Escherichia
 RT coli.*;
 RT submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RN INVOLVEMENT IN KNOBLOCH SYNDROME.
 RP MEDLINE-20400145; PubMed-10942434;
 RX Sertle A.U., Sossi V., Camargo A.A., Zatz M., Brahe C.,
 RA Passos-Bueno M.R.;
 RA *Collagen XVIII, containing an endogenous inhibitor of angiogenesis
 RT and tumor growth, plays a critical role in the maintenance of retinal
 RT structure and in neural tube closure.*;
 RT Hum. Mol. Genet. 9:2051-2058(2000)
 RL [1]
 CC -1- FUNCTION: COL1A18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
 CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULFATE PROTEOGLYCANS INVOLVED IN GROWTH
 CC FACTOR SIGNALING (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOPFORMS: A SHORT FORM/NC1-103 AND A LONG
 CC FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING
 CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
 CC IN LIVER, LUNG AND KIDNEY.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- DISEASE: DEFECTS IN COL18A1 ARE A CAUSE OF KNOBLOCH SYNDROME (KS);
 CC AN AUTOSOMAL RECESSIVE DISORDER DEFINED BY THE OCCURRENCE OF HIGH
 CC MYOPIA, VITREORETINAL DEGENERATION WITH RETINAL DETACHMENT,
 CC MACULAR ABNORMALITIES AND OCCIPITAL ENCEPHALOCELE.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF018081; AAC39658.1; -;
 DR EMBL: AF018082; AAC39659.1; -;
 DR EMBL: AL163302; CAB90482.1; -;
 DR EMBL: L22548; AAA51864.1; -;
 DR EMBL: AF184060; AAF0310.1; ALT_INIT.
 DR GLOSULEDB: P39060; -;
 DR MIM: 120328; -;
 DR MIM: 267750; -;
 DR InterPro: IPR000087; -;
 DR Pfam: PF01391; Collagen: 6.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 1516
 FT CHAIN 1334 1516
 FT CHAIN 24 1516
 FT CHAIN 516 550
 FT CHAIN 516 550
 FT CHAIN 561 640
 FT CHAIN 641 664
 FT CHAIN 665 786
 FT CHAIN 787 809
 FT CHAIN 810 892
 FT CHAIN 810 892

FT DOMAIN 893 906
 FT DOMAIN 907 948
 FT DOMAIN 949 961
 FT DOMAIN 962 1034
 FT DOMAIN 1035 1044
 FT DOMAIN 1045 1077
 FT DOMAIN 1078 1089
 FT DOMAIN 1090 1111
 FT DOMAIN 1112 1118
 FT DOMAIN 1119 1173
 FT DOMAIN 1174 1186
 FT DOMAIN 1187 1204
 FT DOMAIN 1205 1516
 FT CARBOHYD 68 68
 FT CARBOHYD 129 129
 FT CARBOHYD 164 164
 FT CARBOHYD 691 691
 FT DISULFID 1366 1506
 FT DISULFID 1468 1498
 FT SITE 1095 1097
 FT VARSPIC 1 180
 FT VARSPIC 181 215
 FT CONFLICT 428 428
 FT CONFLICT 841 841
 FT CONFLICT 877 877
 FT CONFLICT 886 886
 FT CONFLICT 912 912
 FT CONFLICT 933 933
 FT CONFLICT 975 975
 FT CONFLICT 1064 1064
 FT CONFLICT 1084 1084
 FT CONFLICT 1120 1120
 FT CONFLICT 1123 1123
 FT CONFLICT 1126 1126
 FT CONFLICT 1206 1206
 FT CONFLICT 1304 1304
 FT CONFLICT 1314 1314
 FT CONFLICT 1323 1324
 FT CONFLICT 1437 1437
 FT CONFLICT 1443 1443
 FT CONFLICT 1483 1483
 FT SEQUENCE 1516 AA; 153840 MW; 3C70F29AA4476BE76 CRC64;
 Query Match 92.7%; Score 38; DB 1; Length 1516;
 Best Local Similarity 87.5%; Pred. No. 6.7;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYTYCTE 8
 DB 1501 AYTYCTE 1508
 RESULT 3
 YAI17_HUMAN STANDARD; PRT; 1015 AA.
 ID YAI17_HUMAN
 AC Q90P23;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN KIAA1017.
 GN KIAA1017.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eulheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCB Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
CC -1- SIMILARITY: TO HUMAN KIAA0297/KIAA0329.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB023234; BAA76861.1; -
DR Hypothetical protein.
KW
SQ SEQUENCE 1015 AA; 114842 MW; 021AC2BFD67EA426 CRC64;

```

Query Match      78.0%; Score 32; DB 1; Length 1015;
Best Local Similarity 57.1%; Pred. No. 70;
Matches 4; Conservative 3; Mismatches 0; Gaps 0;

OY      2 YVLCAT 8
         |::|::|
db      875 YVLCLE 881

```

DB 875 YLICLE 881

RESULT	4	
YA36_METZA		
ID	YA36_METZA	STANDARD: PRT; 233 AA.
AC	O58442;	
DT	15-JUL-1998 (rel. 36, Created)	
DT	15-JUL-1998 (rel. 36, Last sequence update)	
DT	15-JUL-1998 (rel. 36, Last annotation update)	
DE	HYPOTHEICAL PROTEIN MJ1036.	
GN	MJ1036.	
OS	Methanococcus jannaschii.	
OC	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;	
OC	Methanococcus.	
OX	NCBI_TaxID=2190;	
RP	(1)	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;	
RC	MEDLINE=66337999; PubMed=6686087;	
RA	Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,	
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,	
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Belch C.I.,	
RA	Overbeek R., Kirnness E.F., Weissstock K.G., Merrick J.W., Glöckner A.,	
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,	
RA	Uttelback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,	
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,	
RA	Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;	
RT	*Complete genome sequence of the methanogenic archaeon, Methanococcus	
TT	jannaschii".	

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/isb-sib.ch>) or send an email to license@isb-sib.ch.

DR	EMBL; U67546; AAB9945.1; -	TIGR; MJ1036; -	Hypothetical protein; Transmembrane.
KW	TRANSMEM	7	POTENTIAL
FT	TRANSMEM	36	POTENTIAL
FT	TRANSMEM	62	POTENTIAL
FT	TRANSMEM	82	POTENTIAL
FT	TRANSMEM	119	POTENTIAL

FT	TRANSMEM	159	179	POTENTIAL.
FT	TRANSMEM	188	208	POTENTIAL.
SO	SEQUENCE	233 AA;	26170 MW;	D2D82ACF6200DB58 CRC64;

Query Match	73.28;	Score 30;	DB 1;	Length 233;
Best Local	Similarity 71.48;	Pred. No. 47;		
Matches	5; Conservative	2; Mismatches	0; Indels	0; Gaps

QY	1	SYIVLCI	7
		1:11:11	
Db	15	SFIVMCI	21

RESULT	5	
RDS_FELCA		
ID	RDS_FELCA	STANDARD: PRT: 346 AA.
AC	P35906;	
DT	01-JUN-1994 (Rel. 29, Created)	
DT	01-NOV-1995 (Rel. 32, Last sequence update)	
DT	01-OCT-1996 (Rel. 34, Last annotation update)	
DE	PERIPHERIN (RETINAL DEGENERATION SLOW PROTEIN).	
GN	RDS.	
OS	Felis silvestris catus (Cat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.	
OX	NCBI_Taxid=9685;	

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/at-license@sib-sib.ch>) or send an email to license@sib-sib.ch.

DR	EMBL: M3404/; AAA191/5.1; -	
DR	InterPro: IPR000830; -	
DR	Pfam: PF00824; RDS_ROM1; 1.	
DR	PRINTS: PRO0218; PERIPHERNRPS.	
DR	PROSITE: PS00930; RDS_ROM1; 1.	
KM	Photoreceptor; Vision; Transmembrane; Glycoprotein.	
FT	DOMAIN 1 24	CTOPLASMIC (POTENTIAL).
FT	TRANSMEM 25 43	POTENTIAL.
FT	DOMAIN 44 61	LUMENAL (POTENTIAL).
FT	TRANSMEM 62 80	POTENTIAL.
FT	DOMAIN 81 99	CTOPLASMIC (POTENTIAL).
FT	TRANSMEM 100 123	POTENTIAL.
FT	DOMAIN 124 264	LUMENAL (POTENTIAL).
FT	TRANSMEM 265 290	POTENTIAL.
FT	DOMAIN 291 346	CTOPLASMIC (POTENTIAL).
FT	CARBOHYD 33 53	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 229 259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE 346 AA: 39171 MW: 4700FOE29E4759A6 CRC64:	

Query Match 73.2%; Score 30; DB 1; Length 346;
 Best Local Similarity 57.18; Pred. No. 67;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 111111
 DB 100 SYIVLCV 106

RESULT 6
 RFC_SALTY STANDARD; PRT; 407 AA.
 ID P26479;
 AC P26479;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE O-ANTIGEN POLYMERASE.
 GN RFC.
 OS Salmonella typhimurium.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmomella.
 CC NCBI_Taxid=602;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE-9193209; PubMed-1707412;
 CC Collins L.V., Hackett J.;
 CC "Molecular cloning, characterization, and nucleotide sequence of the
 CC rfc gene, which encodes an O-antigen polymerase of Salmonella
 CC typhimurium.";
 CC J. Bacteriol. 173:2521-2529(1991).
 CC -1- FUNCTION: MAY LINK THE O-ANTIGEN TETRASACCHARIDE UNITS INTO LONG
 CC CHAINS, GIVING RISE TO TYPICAL SMOOTH LPS.
 CC -1- PATHWAY: LIPIDPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC CONTAINS SEVERAL POTENTIAL TRANSMEMBRANE DOMAINS (POTENTIAL).
 CC -1- SIMILARITY: TO OTHER O-ANTIGEN POLYMERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M60066; AAA27210.1; -;
 CC PIR; A43672; A43672.
 CC StyGene; SG10352; rfc.
 CC LIPIDPOLYSACCHARIDE biosynthesis; Transmembrane; Inner membrane.
 CC KW TRANSMEM 2 22 POTENTIAL.
 CC FT TRANSMEM 31 51 POTENTIAL.
 CC FT TRANSMEM 63 83 POTENTIAL.
 CC FT TRANSMEM 101 121 POTENTIAL.
 CC FT TRANSMEM 141 161 POTENTIAL.
 CC FT TRANSMEM 179 199 POTENTIAL.
 CC FT TRANSMEM 211 231 POTENTIAL.
 CC FT TRANSMEM 320 340 POTENTIAL.
 CC FT TRANSMEM 356 376 POTENTIAL.
 CC FT TRANSMEM 382 402 POTENTIAL.
 CC SEQUENCE 407 AA; 47461 MW; AF187D8633D9CEDE CRC64;

Query Match 73.2%; Score 30; DB 1; Length 407;
 Best Local Similarity 71.4%; Pred. No. 77;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 111111
 DB 5 SYIVLCI 11

RESULT 7
 NAPI_HUMAN STANDARD; PRT; 420 AA.
 ID NAPI_HUMAN
 AC O96009;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NAPSIN 1 PRECURSOR (EC 3.4.23.-) (NAPSIN A) (NAPA) (TA01/TA02)
 DE (ASPARTYL PROTEASE 4) (ASP 4) (ASP4).
 GN NAPI.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_Taxid=9606;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC TISSUE=Lung, and Kidney;
 CC MEDLINE-9902989; PubMed-9877162;
 CC Tainelli P.J., Powell D.J., Hill J., Smith T.S., Tew D.G., Kay J.;
 CC "Napsins: new human aspartic proteases. Distinction between two
 CC closely related genes.";
 CC FEBS Lett. 441:43-48(1998).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC TISSUE=Liver;
 CC Koelsch G., Wu S., Henthorn J., Tang J., Lin X.;
 CC "New human aspartic proteases napsin 1 and napsin 2: molecular
 CC cloning and intracellular localization of napsin 1.";
 CC submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MASS SPECTROMETRY.
 CC TISSUE=Fetal lung;
 CC MEDLINE-2004/840; PubMed-10580105;
 CC Chuman Y., Bergman A.-C., Ueno T., Saito S., Sakaguchi K.,
 CC Alaiya A.A., Franzen B., Bergman T., Arnott D., Auer G., Appella E.,
 CC Joernvall H., Linder S.;
 CC "Napsin A, a member of the aspartic protease family, is abundantly
 CC expressed in normal lung and kidney tissue and is expressed in lung
 CC adenocarcinomas.";
 CC FEBS Lett. 462:129-134(1999).
 CC [4]
 CC SEQUENCE FROM N.A.
 CC Blankowski M.J., Shuck M.E., Slightom J.L., Drong R.P.;
 CC submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PROCESSING OF PNEUMOCYTE SURFACTANT
 CC PRECURSORS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ADULT LUNG (TYPE II
 CC PNEUMOCYTES) AND KIDNEY AND IN FETAL LUNG. LOW LEVELS IN ADULT
 CC SPLEEN AND VERY LOW LEVELS IN PERIPHERAL BLOOD LEUKOCYTES
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EURKARTOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF090386; AAD04917.1; -;
 CC EMBL; AF098484; AAD13215.1; -;
 CC EMBL; AF200345; AAF17081.1; -;
 CC DR MEROPS; A01.046; -;
 CC DR InterPro; IPR001461; -;
 CC DR Pfam; PF00026; asp. 1.
 CC DR PRINTS; PR00792; PEPSIN.
 CC DR PROSITE; PS00141; ASP_PROTEASE; 2.
 CC KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 CC FT SIGNAL 1 25 POTENTIAL.
 CC FT PROPEP 26 63 ACTIVATION PEPTIDE.
 CC FT CHAIN 64 420 NAPSIN 1.

FT ACT_SITE 96 96 BY SIMILARITY.
 FT ACT_SITE 283 283 BY SIMILARITY.
 FT DISULFID 109 116 BY SIMILARITY.
 FT DISULFID 274 278 BY SIMILARITY.
 FT DISULFID 317 354 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 420 AA; 45386 MW; 018B86AE5BD0C865 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 420;
 Best Local Similarity 71.4%; Pred. No. 79;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YIVLCIE 8
 Db 313 YIILCE 319

RESULT 8
 ID NP73_HUMAN STANDARD; PRT: 436 AA.
 AC 000624;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
 DE COTRANSPORTER 3) (NA(+)/PI COTRANSPORTER 3).
 GN SLC17A3 OR NP73.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ruddy D.A., Krommal G.S., Lee V.K., Mintler G.A., Quintana L.,
 RA Domingo R., Meyer N.C., Basava A., McClelland E., Fullan A.,
 RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
 RA Wolff R.K., Schatzman R.C., Feder J.N.,
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
 CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U91328; AAB82085.1; -
 DR EMBL: U90544; AAB53422.1; -
 KM Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
 KW Sodium transport.
 FT TRANSMEM 36 46 POTENTIAL.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 130 150 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 317 337 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 436 AA; 47350 MW; DF02F618B3A572F CRC64;

Query Match 73.2%; Score 30; DB 1; Length 436;

Best Local Similarity 57.1%; Pred. No. 82;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYIVLCI 7
 Db 193 SFYILCV 199

RESULT 9
 ID YH40_YEAST STANDARD; PRT: 550 AA.
 AC 003212; Q03830;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 62.5 KDA PROTEIN IN ALD2-DDP48 INTERGENIC REGION.
 GN YMR171C OR YH6010.01C OR YH6520.20C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-241 FROM N.A.
 RA STRAIN-S288C / AB972;
 RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 215-550 FROM N.A.
 RA STRAIN-S288C / AB972;
 RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.,
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO YEAST YKL124W.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 249808; CAA89904.1; -
 DR EMBL: 249705; CAA89807.1; -
 DR SGD: S0004781; YMR171C.
 DR InterPro: IPR000107; -
 DR Pfam: PF00622; Spry; 1;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 46 66 POTENTIAL.
 SO SEQUENCE 550 AA; 62532 MW; 6B1C6E9F9A89FA CRC64;

Query Match 73.2%; Score 30; DB 1; Length 550;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YIVLCI 7
 Db 58 YIVICI 63

RESULT 10
 ID COMP_BACSU STANDARD; PRT: 769 AA.
 AC 099027; Q05226;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SENSOR PROTEIN COMP (EC 2.7.1.-).
 GN COMP.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;

```

[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BD630:
RA MEDLINE-90337321: PubMed-2116363;
RA Weinrauch Y., Penchev R., Dubnau D.;
RT "A Bacillus subtilis regulatory gene product for genetic competence
RT and sporulation resembles sensor protein members of the bacterial
RT two-component signal-transduction systems.";
RL Genes Dev. 4:860-872(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Oudega B., Koningssteijn G., Van Zoest A.;
RT "Bacillus subtilis genome project, DNA sequence from yufA to
RT yufD ";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-16 FROM N.A.
RC STRAIN-168;
RA MEDLINE-91358358: PubMed-1715859;
RA Weinrauch Y., Msadek T., Kunst F., Dubnau D.;
RT "Sequence and properties of comP, a new competence regulatory gene of
RT Bacillus subtilis.";
RL J. Bacteriol. 173:5685-5693(1991).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM COMP/COMA.
CC HISTIDINE KINASE THAT IS REQUIRED EARLY IN THE COMPETENCE CASCADE.
CC ACTIVATES COMA PROTEIN BY PHOSPHORYLATION. IT PLAYS ROLE IN
CC SPOREATION. AT LEAST PARTLY INTERCHANGEABLE WITH THAT OF SPOIIJ.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC REGULATORY PROTEINS WHICH BELONG
CC TO A TWO-COMPONENT REGULATORY SYSTEM AND TRANSDUCE ENVIRONMENTAL
CC SIGNALS TO TRANSCRIPTIONAL APPARATUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X54010; CA37957.1; -
DR EMBL: 293932; CAB07903.1; -
DR EMBL: M22856; AAA22319.1; -
DR EMBL: M71283; AAA22324.1; -
DR EMBL: 299120; CAB15157.1; -
DR PIR: A35848; A35848.
DR Subtilisin; BG10380; comp.
KW Sensory transduction; Phosphorylation; Transferase; Kinase;
KW Transmembrane.
FT DOMAIN 1 9
FT TRANSMEM 10 33
FT TRANSMEM 34 113
FT TRANSMEM 114 134
FT TRANSMEM 135 144
FT TRANSMEM 145 167
FT TRANSMEM 168 235
FT TRANSMEM 236 257
FT TRANSMEM 258 272
FT TRANSMEM 273 295
FT TRANSMEM 296 299
FT TRANSMEM 300 323
FT TRANSMEM 324 337
FT TRANSMEM 338 357
FT TRANSMEM 358 361
FT TRANSMEM 362 383
FT TRANSMEM 384 769
FT DOMAIN 533 769
FT MOD_RES 456 456
FT CONFLICT 604 604
FT CONFLICT 610 610
FT CONFLICT 628 628
FT CONFLICT 636 637

```

```

SQ SEQUENCE 769 AA: 89316 MW: 5528278 DE5B4 CRC64;
Query Match
Best Local Similarity 73.2%; Score 30; DB 1; Length 769;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYVLCT 7
Db 497 SYVLCT 503
RESULT 11
TH1L_SCHPO
ID TH1L_SCHPO STANDARD; PRT; 775 AA.
AC P36598; P40378;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE THIAMINE REPRESSIBLE GENES REGULATORY PROTEIN TH1L (TRANSCRIPTION
DE FACTOR NTF1).
GN TH1L OR NTF1.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA MEDLINE-94374699; PubMed-8088540;
RA Frankhauser H., Schweingruber M.E.;
RT "Thiamine-repressible genes in Schizosaccharomyces pombe are
RT regulated by a Cys6 zinc-finger motif-containing protein.";
RL Gene 147:141-144(1994).
RN [2]
RP SEQUENCE OF 1-728 FROM N.A.
RC STRAIN-972;
RA MEDLINE-94216297; PubMed-8163491;
RA Tang C.S.L., Bueno A., Russell P.;
RT "ntf1 encodes a 6-cysteine zinc finger-containing transcription
RT factor that regulates the nmt1 promoter in fission yeast.";
RL J. Biol. Chem. 269:11921-11926(1994).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT ACTIVATES THE NMT1 PROMOTER.
CC REGULATION OF THIAMINE REPRESSIBLE GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77512; CAAS4648.1; -
DR EMBL: L25912; AAA19010.1; -
DR PIR: S41962; S41962.
DR PIR: S43749; S43749.
DR PIR: A53575; A53575.
DR HSSP: P04386; ID66.
DR InterPro: IPR001138; -.
DR Pfam: PF001172; Zn_c1us; 1.
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE: PS50048; ZN2_Cy6_FUNGAL_2; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Zinc_Metal-binding.
FT DNA_BIND 39 65
FT CONFLICT 10 10
FT CONFLICT 325 326
FT CONFLICT 684 684
S -> C (IN REF. 2).
S -> K (IN REF. 2).
S -> LR (IN REF. 2).
S -> C (IN REF. 2).

```

SEQUENCE 775 AA: 88060 MW: 88E2A48072F678FB CRC64:

Query Match 73.2%; Score 30; DB 1; Length 775;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2 YIVLCIE 8
 547 YIVLCVE 553

RESULT 12
 ID VP_H1_NEUCR STANDARD: PRT: 856 AA.
 AC 001290;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VACUOLAR ATP SYNTHASE 98 KDA SUBUNIT (EC 3.6.1.34) (VACUOLAR ATPASE 98 KDA SUBUNIT).
 GN VP_H1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_Taxid=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A;
 RA Bowman E.J.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR
 ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF
 THE ENZYME FOR A SPECIFIC ORGANELLE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR.
 CC -1- SIMILARITY: BELONGS TO THE VAPASE 116 KDA SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb.ch).
 CC -----
 DR EMBL: U36396; AAA93078.1; -
 DR InterPro: IPR002430; -
 DR Pfam: PF01496; V_Atpase_sub_a; 1.
 KW Hydrogen ion transport; Transmembrane; Glycoprotein.
 FT DOMAIN 1 415 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 416 436 POTENTIAL.
 FT DOMAIN 437 457 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 458 478 POTENTIAL.
 FT DOMAIN 479 548 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 549 569 POTENTIAL.
 FT DOMAIN 570 581 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 582 602 POTENTIAL.
 FT DOMAIN 603 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 636 POTENTIAL.
 FT DOMAIN 637 642 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 643 663 POTENTIAL.
 FT DOMAIN 664 754 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 755 775 POTENTIAL.
 FT DOMAIN 776 786 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 787 807 POTENTIAL.
 FT DOMAIN 808 856 EXTRACELLULAR (POTENTIAL).
 FT CAROXYD 856 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 856 AA: 97992 MW: F75E3737A7AD3DE8 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 856;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

2 YIVLCI 7
 595 YIVLCI 600

RESULT 13
 ID CAIE_HUMAN STANDARD: PRT: 1388 AA.
 AC P39059;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(XV) CHAIN PRECURSOR.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heilmaki P., Reim M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 exon-intron organization in the 3' region of the corresponding
 gene";
 RT J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Murgaki Y., Abe N., Niinomiya Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 non-triple helical domain with a tandem repeat structure and homology
 to alpha 1(XVII) collagen";
 RT J. Biol. Chem. 269:4042-4046(1994).
 RN [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 1(XV), characterized by extensive interruptions in the triple-helical
 region";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
 SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb.ch).
 CC -----
 DR EMBL: U25286; AAA58429.1; -
 DR EMBL: D21230; BAA04762.1; -
 DR EMBL: L01697; -; NOT_ANNOTATED_CDS.
 DR MIM: 120325; -
 DR InterPro: IPR000087; -
 DR Pfam: PF01391; Collagen; 4.
 KW Cell adhesion; Collagen; Glycoprotein; Signal.
 FT CHAIN 1 25 POTENTIAL.
 FT SIGNAL 26 1388 COLLAGEN ALPHA 1(XV) CHAIN.
 FT DOMAIN 26 555 NONHELICAL REGION 1 (NC1).
 FT DOMAIN 556 573 TRIPLE-HELICAL REGION 1 (COL1).
 FT DOMAIN 574 618 NONHELICAL REGION 2 (NC2).
 FT DOMAIN 619 732 TRIPLE-HELICAL REGION 2 (COL2).
 FT


```

FT DOMAIN 733 763 NONHELICAL REGION 3 (NC3).
FT DOMAIN 764 798 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 799 822 TRIPLE-HELICAL REGION 4 (NC4).
FT DOMAIN 823 867 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 868 878 TRIPLE-HELICAL REGION 5 (NC5).
FT DOMAIN 879 949 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 950 983 TRIPLE-HELICAL REGION 6 (NC6).
FT DOMAIN 984 1013 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1014 1027 TRIPLE-HELICAL REGION 7 (NC7).
FT DOMAIN 1028 1045 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1046 1052 TRIPLE-HELICAL REGION 8 (NC8).
FT DOMAIN 1053 1107 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1108 1117 TRIPLE-HELICAL REGION 9 (NC9).
FT DOMAIN 1118 1132 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1133 1388 NONHELICAL REGION 10 (NC10).
FT DOMAIN 358 408 4 X TANDEM REPEATS.
FT REPEAT 409 459 1.
FT REPEAT 460 459 2.
FT REPEAT 509 509 3.
FT REPEAT 510 555 4.
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 687 687 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 10 10 C -> S (IN REF. 2).
FT CONFLICT 49 49 D -> V (IN REF. 2).
FT CONFLICT 95 95 L -> A (IN REF. 2).
FT CONFLICT 150 150 P -> A (IN REF. 2).
FT CONFLICT 204 204 M -> V (IN REF. 2).
FT CONFLICT 409 409 R -> A (IN REF. 2).
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;

```

```

Query Match          73.2%; Score 30; DB 1; Length 1388;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 IVLCIE 8
Db 1374 IVLCIE 1379

```

```

RESULT 14
YOTO_CAEEL
ID YOTO_CAEEL STANDARD; PRT; 80 AA.
AC P34655;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOPHETICAL 8.7 KDA PROTEIN ZK632.10 IN CHROMOSOME III.
GN ZK632.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6233;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson J., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton L., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen M.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Stilson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans.*
RL Nature 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0057 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22181; CA80190.1; -.
DR PIR; S40942; S40942.
DR WormPep; ZK632.10; CE02385.
DR InterPro; IPR000612; -.
DR Pfam; PF01679; UPF0057.1.
DR PROSITE; PS01309; UPF0057.1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
SQ SEQUENCE 80 AA; 8708 MW; 3741ABE9B2DA1A33 CRC64;

```

```

Query Match          70.7%; Score 29; DB 1; Length 80;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 YIVLCIE 8
Db 48 YIVLCIE 54

```

```

RESULT 15
MCSR_HUMAN
ID MCSR_HUMAN STANDARD; PRT; 325 AA.
AC P33032;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MELANOCORTIN-5 RECEPTOR (MCS-R) (MC-2).
GN MCSR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE=93384614; PubMed=8396929;
RA Chhajlani V., Muceniece R., Wikberg J.E.S.;
RT "Molecular cloning of a novel human melanocortin receptor.*";
RL Biochem. Biophys. Res. Commun. 195:866-873(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94234987; PubMed=8179577;
RA Griffin N., Mignon V., Facchinetti P., Diaz J., Schwartz J.C.,
RA Sokoloff P.;
RT "Molecular cloning and characterization of the rat fifth melanocortin
RT receptor.*";
RL Biochem. Biophys. Res. Commun. 200:1007-1014(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE=95258173; PubMed=7739752;
RA Fath Z., Iben L.G., Parker E.M.;
RT "Cloning, expression, and tissue distribution of a fifth melanocortin
RT receptor subtype.*";
RL Neurochem. Res. 20:107-113(1995).
CC -1- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE
CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC ADENYLATE CYCLASE. THIS RECEPTOR IS A POSSIBLE MEDIATOR OF THE

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:38:57 ; Search time 12.81 Seconds
(without alignments)
47.572 Million cell updates/sec

Title: US-09-589-777a-2_COPY_168_175

Perfect score: 41

Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.68.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1315	2 A56101	collagen alpha 1(X
2	41	100.0	1774	2 B56101	collagen alpha 1(X
3	38	92.7	684	2 A53019	collagen alpha 1(X
4	32	78.0	147	2 F71498	hypothetical prote
5	32	78.0	147	2 E81656	conserved hypotet
6	32	78.0	293	2 D96505	unknown protein fl
7	32	78.0	1142	2 A45031	cysteine-rich fibr
8	32	78.0	1175	2 S52417	E-selectin ligand
9	31	75.6	39	2 G85837	hypothetical prote
10	31	75.6	158	2 H72029	conserved hypotet
11	31	75.6	158	2 D86593	hypothetical prote
12	31	75.6	163	2 D84853	hypothetical prote
13	31	75.6	235	2 T23501	hypothetical prote
14	31	75.6	316	2 B75323	conserved hypotet
15	31	75.6	466	2 B82482	alpha-amylase VCAO
16	31	75.6	479	2 F86285	hypothetical prote
17	31	75.6	566	2 E81404	acetylactate synth
18	31	75.6	2342	2 T18200	fatty-acid synthas
19	30	73.2	96	2 T42273	hypothetical prote
20	30	73.2	129	2 T21687	hypothetical prote
21	30	73.2	141	2 T16105	hypothetical prote
22	30	73.2	151	2 T48823	clathrin coat asse
23	30	73.2	163	2 T33130	hypothetical prote
24	30	73.2	168	2 T12834	hypothetical prote
25	30	73.2	233	2 C64429	hypothetical prote
26	30	73.2	237	2 D83313	probable permease
27	30	73.2	277	2 D84596	hypothetical prote
28	30	73.2	346	2 I46087	periphetin - cat
29	30	73.2	353	2 G71567	probable cytochrom

30	30	73.2	407	2 A43672	O-antigen polymera
31	30	73.2	507	2 D86207	hypothetical prote
32	30	73.2	550	2 S55118	probable membrane
33	30	73.2	769	2 A35848	competence regulat
34	30	73.2	769	2 B69604	two-component sens
35	30	73.2	772	2 T52402	hypothetical prote
36	30	73.2	775	2 S41962	thiamin repressibl
37	30	73.2	947	2 T01238	hypothetical prote
38	30	73.2	1206	2 E86445	hypothetical prote
39	30	73.2	1388	2 A53317	collagen alpha 1(X
40	29	70.7	70	2 S54439	hemstruv operon pro
41	29	70.7	76	2 T24688	hypothetical prote
42	29	70.7	92	2 S40942	hypothetical prote
43	29	70.7	93	2 F45252	pliv variable regl
44	29	70.7	145	2 T11048	hypothetical prote
45	29	70.7	187	2 T09390	21k protein precur

ALIGNMENTS

RESULT 1
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N:Contains: endostatin
C:Species: Mus musculus (house mouse)
C>Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000
C/Accession: A56101; A58371; S72450; S65595
R:Rehn, M.; Philajantemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A>Title: Identification of three N-terminal ends of type XVIII collagen chains and t
lif homologous to rat and Drosophila frizzled proteins.
A:Reference number: A56101; M01D:95181468
A:Accession: A56101
A:Molecule type: mRNA
A:Residues: 1-103 <REH1>
A:Cross-references: GB:U11636; NID:9618427; PIDN:AA052178.1; PID:9618428
R:Rehn, M.; Philajantemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A>Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno
A:Reference number: A58371; M01D:94240112
A:Accession: A58371
A:Molecule type: mRNA
A:Residues: 1-928 <REH2>
A:Cross-references: GB:U116998; NID:9404754; PIDN:AAA7434.1; PID:9553894
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
Submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 28-687, 'U', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A>Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-
A:Reference number: A58370; M01D:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 28-1315 <OHS>
A:Cross-references: EMBL:L22545
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:Col18a1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-335/Region: thrombospondin amino-terminal similarity

```

F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M>
F:327-353/Domains: collagenous #status predicted <CO1>
F:364-437/Domains: collagenous #status predicted <CO2>
F:462-583/Domains: collagenous #status predicted <CO3>
F:607-688/Domains: collagenous #status predicted <CO4>
F:704-743/Domains: collagenous #status predicted <CO5>
F:759-831/Domains: collagenous #status predicted <CO6>
F:842-874/Domains: collagenous #status predicted <CO7>
F:887-910/Domains: collagenous #status predicted <CO8>
F:892-894/Region: cell attachment (R-G-D) motif
F:918-966/Domains: collagenous #status predicted <CO9>
F:983-1000/Domains: collagenous #status predicted <CO10>
F:1133-1315/Product: endostatin #status predicted <EST>
F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
F:126-488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:172-228/Dissulfide bonds: #status predicted
F:240-245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match      100.0%; Score 41; DB 2; Length 1315;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYTVLCIE 8
Db 1299 SYTVLCIE 1306

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
M:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000
C:Accession: B56101; C56101; S72450; S65595; PNO675; A54072; A58816
R:Rehn, M.; Philantaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and their
clft homologous to rat and Drosophila frizzled proteins.
A:Reference number: A56101; MUID:95181468
A:Accession: B56101
A:Molecule type: mRNA
A:Residues: 1-562 <REH>
A:Cross-references: GB:U11637; NID:9618429; PIDN:AA52179.1; PID:9618430
A:Experimental source: splice form clone PE17.24
A:Accession: C56101
A:Molecule type: mRNA
A:Residues: 1-239,487-562 <REH2>
A:Cross-references: GB:U11637; NID:9618429
A:Experimental source: splice form clones PE8.1, PE19, PE15.2
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A:Reference number: A58370; MUID:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A:Cross-references: EMBL:L22545
R:Rehn, M.; Muragaki, Y.; Yoshioaka, H.; Olsen, H.; Nishimura, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A:Title: Identification of a novel collagen chain represented by extensive interruptions
A:Reference number: PNO675; MUID:94059075
A:Accession: PNO675
A:Molecule type: mRNA
A:Residues: 635-1774 <ABE>

```

```

R:Rehn, M.; Hunkler, E.; Philantaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial
collagen chain.
A:Reference number: A54072; MUID:94245707
A:Accession: A54072
A:Molecule type: DNA; mRNA
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
A:Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734
R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.;
Cell 88, 277-285, 1997
A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A:Reference number: A58816; MUID:97160848
A:Accession: A58816
A:Molecule type: protein
A:Residues: 1591-1610 <ORE>
A:Experimental source: hemangioendothelium cells
A:Note: Inhibits endothelial cell proliferation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:Coll8a1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15
A:Note: the list of introns is incomplete
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; collagen alpha 1(XVIII) chain precursor, long splice form
F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
F:1-24/Domains: signal sequence #status predicted <SIG>
F:361-486/Region: frizzled similarity
F:786-812/Domains: collagenous #status predicted <CO01>
F:823-896/Domains: collagenous #status predicted <CO02>
F:921-1042/Domains: collagenous #status predicted <CO03>
F:1066-1148/Domains: collagenous #status predicted <CO04>
F:1166-1204/Domains: collagenous #status predicted <CO05>
F:1218-1290/Domains: collagenous #status predicted <CO06>
F:1301-1333/Domains: collagenous #status predicted <CO07>
F:1346-1369/Domains: collagenous #status predicted <CO08>
F:1351-1353/Region: cell attachment (R-G-D) motif
F:1377-1428/Domains: collagenous #status predicted <CO09>
F:1442-1459/Domains: collagenous #status predicted <CO10>
F:1591-1774/Product: endostatin #status predicted <EST>
F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match      100.0%; Score 41; DB 2; Length 1774;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYTVLCIE 8
Db 1758 SYTVLCIE 1765

RESULT 3
A53019
collagen alpha 1(XVIII) chain - human (fragment)
M:Contains: endostatin
C:Species: Homo sapiens (man)
C:Date: 18-Aug-1995 #sequence-revision 18-Aug-1995 #text-change 31-Mar-2000
C:Accession: A53019
R:Oh, S.P.; Watanabe, M.L.; Seidlin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse
Genomics 19, 494-499, 1994
A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local
A:Reference number: A53019; MUID:94245237

```

A:Accession: A53019
 A:Molecule type: mRNA
 A:Residues: 1-684 <ORF>
 A:Cross-references: GB:L22548; NID:9348908; PIDN:AA51864.1; PID:9562794
 A:Note: The cited accession number, L22548, is not in Genbank release 103
 A:Note: In the authors' translation, 482-614 is not shown, residues 483-490 are shifted
 C:Comment: Prolins and lysines at the third position of the tripeptide repeating unit
 C:Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
 C:Genetics: GDB:COL18A1
 A:Gene: GDB:COL18A1
 A:Cross-references: GDB:138752; OMIM:120328
 A:Map position: 21q22.3-21q22.3
 C:Superfamily: unassigned collagens
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
 F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
 F:1-59/Domain: collagenous (fragment) #status predicted <CO5>
 F:74-115/Domain: collagenous #status predicted <CO4>
 F:129-201/Domain: collagenous #status predicted <CO6>
 F:212-244/Domain: collagenous #status predicted <CO7>
 F:257-278/Domain: collagenous #status predicted <CO8>
 F:262-264/Region: cell attachment (R-G-D) motif
 F:286-340/Domain: collagenous #status predicted <CO9>
 F:354-371/Domain: collagenous #status predicted <CO10>
 F:502-684/Product: endostatin #status predicted <EST>
 F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 92.7%; Score 38; DB 2; Length 684;
 Best Local Similarity 87.5%; Pred. No. 6.5;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 8
 Db 669 AYIVLCIE 676

RESULT 4
 F71498
 hypothetical protein CT565 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: F71498
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 287, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
 A:Reference number: A71570; MUID:99000809
 A:Accession: F71498
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <ARN>
 A:Cross-references: GB:AE001327; GB:AE001273; NID:93328999; PIDN:AAC68167.1; PID:9332900
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: CT565

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 123 SFIVLCI 129

RESULT 5
 E81656
 conserved hypothetical protein TC0854 [imported] - Chlamydia muridarum (strain Nigg)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: E81656
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: E81656
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <TEF>
 A:Cross-references: GB:AE002352; GB:AE002160; NID:97190879; PIDN:AAF39650.1; PID:9719
 A:Experimental source: strain Nigg (Mopn)
 C:Genetics:
 A:Gene: TC0854

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 123 SFIVLCI 129

RESULT 6
 D96605
 unknown protein F13N6.19 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96605
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D96605
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-293 <STO>
 A:Cross-references: GB:AE005173; NID:911024843; PIDN:AAG26928.1; GSPDB:GNO0141
 C:Genetics:
 A:Gene: F13N6.19
 A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 293;
 Best Local Similarity 85.7%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 122 SYIALCI 128

RESULT 7
 A45031
 cysteine-rich fibroblast growth factor receptor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A45031
 R:Barut, L.W.; Zuber, M.E.; Lueddecke, B.A.; Olwin, B.B.
 Mol. Cell Biol. 12, 5600-5609, 1992
 A:Title: Identification of a cysteine-rich receptor for fibroblast growth factors.
 A:Reference number: A45031; MUID:93078761
 A:Accession: A45031
 A:Status: preliminary

A:Molecule type: mRNA; protein
A:Residues: 1-1142 <B0R>
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776
A:Experimental source: embryos, brain
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2
C:Keywords: growth factor receptor

Query Match 78.0%; Score 32; DB 2; Length 1142;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
||::||:
Db 362 SYIVLCIE 369

RESULT 8
S52417
E-selectin ligand-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S52417
R:Steegmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl
Nature 373, 615-620, 1995
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa
A:Reference number: S52417; MUID:95157635
A:Accession: S52417
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1175 <STE>
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CA58855.1; PID:g673436
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2

Query Match 78.0%; Score 32; DB 2; Length 1175;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
||::||:
Db 395 SYIVLCIE 402

RESULT 9
G85837
hypothetical protein 23270 [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85837
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; DiMantano, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-39 <STO>
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AA657163.1; GSPDB:GN00145; UMGSP:232
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23270

Query Match 75.6%; Score 31; DB 2; Length 39;
Best Local Similarity 57.1%; Pred. No. 13;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 YIVLCIE 8
::||:::
Db 19 FVIVLCIE 25

RESULT 10
H72029
conserved hypothetical protein CP1049 [Imported] - Chlamydia pneumoniae (strains

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: H72029; A81508
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72029

A:Molecule type: DNA
A:Residues: 1-158 <ARN>
A:Cross-references: GB:AE001663; GB:AE001363; NID:g4377134; PIDN:AA018971.1; PID:g437
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gail, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: A81508
A:Molecule type: DNA
A:Residues: 1-158 <REA>
A:Cross-references: GB:AF002262; GB:AE002161; NID:g7189961; PIDN:AAE38822.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0822; CP1049

Query Match 75.6%; Score 31; DB 2; Length 158;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
||::||:
Db 122 SFIVLCV 128

RESULT 11
D86593
CT9565 hypothetical protein [Imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: D86593
R:Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <STO>
A:Cross-references: GB:BA000008; NID:g8979196; PIDN:BAA99030.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0822

Query Match 75.6%; Score 31; DB 2; Length 158;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
||::||:
Db 122 SFIVLCV 128

RESULT 12
D84853
hypothetical protein At2g42390 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84853
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.;
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Mature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: D84853
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <STO>
A:Cross-references: GB:AE002093; NID:94567311; PIDN:AMD23722.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g42390
A:Map position: 2

Query Match 75.6%; Score 31; DB 2; Length 163;
Best Local Similarity 62.5%; Pred. No. 46;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
11:1111
DB 7 SYLVLCCE 14

RESULT 13
T23501
hypothetical protein K08G2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C:Accession: T23501
R:Lloyd, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: 219749
A:Accession: T23501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-235 <WIL>
A:Cross-references: EMBL:Z81100; PIDN:CAB03191.1; GSPDB:GN00023; CESP:K08G2.2
C:Genetics:
A:Experimental source: clone K08G2
A:Gene: CESP:K08G2.2
A:Map position: 5
A:Intons: 52/2; 92/1; 110/3
C:Superfamily: Caenorhabditis hypothetical protein C49G7.2

Query Match 75.6%; Score 31; DB 2; Length 235;
Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
111111
DB 133 NYIVICI 139

RESULT 14
B75323
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75323
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.; Mc
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: B75323
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-316 <WHI>
A:Cross-references: GB:AE002040; GB:AE000513; NID:96459824; PIDN:AAF11588.1; PID:9645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2041
A:Map position: 1

Query Match 75.6%; Score 31; DB 2; Length 316;
Best Local Similarity 62.5%; Pred. No. 82;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
11:1111
DB 113 SYLVLCFE 120

RESULT 15
B82482
alpha-amylase VCA0250 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82482
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82482
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <HEI>
A:Cross-references: GB:AE004365; GB:AE003853; NID:9657642; PIDN:AAF96161.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0250
A:Map position: 2
C:Superfamily: mammalian alpha-amylase; alpha-amylase core homology

Query Match 75.6%; Score 31; DB 2; Length 466;
Best Local Similarity 83.3%; Pred. No. 12e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLC 6
11:1111
DB 4 SYVVLCC 9

Search completed: August 16, 2001, 11:41:44
Job time: 167 sec

